University of Bath

PHD

Biological and evolutionary significance of cysteine-rich 'B' class pollen coat proteins (PCP-Bs) in the early stages of the Arabidopsis thaliana pollen-stigma interaction

Wang, Ludi

Award date:
2017

Awarding institution:
University of Bath

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 22. May. 2019

Biological and evolutionary significance of cysteine-rich 'B' class pollen coat proteins (PCP-Bs) in the early stages of the *Arabidopsis thaliana* pollen-stigma interaction

Ludi Wang

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

November 2016

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis/portfolio rests with the author and copyright of any previously published materials included may rest with third parties. A copy of this thesis has been supplied on condition that anyone who consults it understands that they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation with effect from…….

Signed on behalf of the Faculty of Science

To my parents, who always support me and encourage me to find a way of life I love, to be brave, to be who I am.

Acknowledgements

I would like to thank my supervisor Dr James Doughty for the advice and support to this project. Dr Susan Crennell for assistance with protein structural predictions, Ursula Potter for support with electron microscopy, Dr Jose Gutierrez-Marcos for the kind gift of *GUS* reporter lines, Prof. Laurence Hurst for the advice on evolutionary analysis, Dr Jean Van Den Elsen for the help on protein purification, Dr Jody Mason for the advice on protein crosslinkning, Prof. Michael Danson for the HPLC equipment, Dr Kate Heesom for the help on mass spectrometry, Dr Shyam Masakapalli and Dr Christopher Ibenegbu for their help on HPLC operating, Dr Francoise Koumanov for the guidance of ultracentrifugation, Emanuele Kendrick on the yeast culture, Prof. Hugh Dickinson, Prof. Simon Hiscock and Prof. Daphne Goring for insightful discussions. I thank Jasmine Zaini and Rebecca Entwistle for valuable contributions during their projects. I thank Prof. Rod Scott, Dr Julia Tratt, Dr Maha Aljabri, Dr Philippe Mozzanega, Dr Alexander Topham, Yaxiao Li, Fan Lian, Dr Jianfeng Xu, Dr Baoxiu Qi and everyone else helped me in the Lab 3 South 1.13. I thank the technician team for lab and equipment maintenance. For the financial support during this PhD project, I would like to thank the University of Bath for the postgraduate scholarship and my parents for their generous help. I thank my partner David Riley for his support and encouragement.

Abbreviation

°C	degree Celsius
A595nm	Absorbance at 595nm
ABA	Abscisic acid
ACA	autoinhibited Ca^{2+}-ATPase
AFP	antifungal protein
AMP	antimicrobial peptides
APS	ammonium persulphate
ARC	Armadillo repeat containing
ARK	Arabidopsis receptor kinase
BAP	basal layer antifungal protein
BEB	Bayes Empirical Bayes
BETL	basal endosperm transfer layer
BiP	Immunoglobulin-binding protein
BLAST	Basic Local Alignment Search Tool
bp	base pair
BSA	bovine serum albumin
Carb	carbenicillin
cDNA	complementary DNA
cer	*eceriferum*
CRP	cysteine-rich protein
CRRSP	cysteine-rich repeat secretory protein
CSα/β	cysteine-stabilised α-helix β-sheet motif
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DEFL	defensin-like protein
DEPC	diethylpyrocarbonate
df	degree of freedom
DNA	deoxyribo-nucleic acid
E.coli	*Escherichia coli*
EAC	Escape from Adaptive Conflict
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EFP	Epidermal Patterning Factors
ENOD	early nodulin
ER	endoplasmic reticulum
ESF	embryo surrounding factor
EXL	extracellular lipase
EXO	exocyst complex
FDR	False discovery rate
GO	gene ontology

GRP	glycine-rich protein
GSI	gametophytic self-incompatibility
GTA	glutaraldehyde
GUS	β-glucuronidase
HPLC	high-performance liquid chromatography
HRP	horseradish peroxidase
IgG	Immunoglobulin G
IMAC	immobilised metal-affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
iTOL	Interactive Tree of Life
kb	kilobase
kDa	kilodalton
LAM	laser-assisted microdissection
LB	lysogeny broth medium
LCR	low-molecular-weight cysteine-rich protein
LRR	leucine-rich repeat
LRT	likelihood ratio test
LTP	lipid transfer protein
MAPK	mitogen-activated-protein kinase
MEG	Maternally Expressed Gene
MEGA	Molecular Evolutionary Genetics analysis
MEME	Multiple Expression Motif for Motif Elicitation
MilliQ (water)	ultrapurified water
MMC	meristemoid mother cell
MP	movement protein
Mr	relative molecular mass
mRNA	messenger RNA
MS	mass spectrophotometry
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NASC	Nottingham Arabidopsis Stock Centre
NCR	nodule-specific cysteine-rich protein
NEB	naive empirical Bayes
NF	nodule factor
NJ	neighbour-joining method
NMR	nuclear magnetic resonance
NO	nitric oxide
nsLTP	non-specific lipid-transfer proteins
OD600	optical density at 600nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAML	Phylogenetic Analysis by Maximum Likelihood
PANTHER	Protein ANalysis THrough Evolutionary Relationships
PBS	phosphate buffered saline
PCP	pollen coat protein
PCR	polymerase chain reaction
PDLP	PD-located protein

plasmodesmata	PD
PRK	pollen-specific receptor-like kinase
PrpS	*Papaver rhoeas* pollen *S*-determinant
PrsS	*Papaver rhoeas* stigma *S*-determinant
PSM	peptide-spectrum match
PVDF	Polyvinylidene difluoride
RACE	Rapid Amplification of cDNA End
RALF	Rapid Alkalinisation Factor
RE	restriction endonuclease (or restriction enzyme)
RFP	red fluorescent protein
RLK	receptor-like kinase
RNA	ribonucleic acid
RNS	reactive nitrogen species
ROS	reactive oxygen species
RPC	reverse phase chromatography
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcripted PCR
SCA	Small Cysteine Adhesion
SCB	Sodium Cacodylate Buffer
SCR	*S*-locus cysteine-rich
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SEM	scanning electron microscope
SI	self-incompatibility
SLF	*S*-locus F-box protein
SLG	*S*-locus glycoprotein
SLGC	stomatal-lineage ground cell
SLR	*S*-locus related
SRK	*S*-receptor kinase
SSI	sporophytic self-incompatibility
strep	streptomycin
TAE	Tris acetate EDTA
Taq	Thermophilus aquaticus (DNA polymerase)
TCEP	Tris (2-carboxyethyl) phosphine
TCP	total cell protein
T-DNA	transfer DNA
TEM	Transmission electron microscopy
TEMED	N,N,N',N' tetramethylethylenediamine
Tris	Tris (hydroxymethyl) methylamine
TTE	transmitting tract epidermis
U	unit (of enzyme activity)
UBL	Ubiquitin-like protein
UI	unilateral incompatibility
v/v	volume per volume

VERL	vitelline envelope receptor for lysin
w/v	weight per volume
WGD	whole genome duplication
Y2H	yeast-two-hybrid
β-ME	β-Mercaptoethanol

Abstract

The early stages of post-pollination in angiosperms involve multiple phases of interaction between male and female reproductive tissues. The establishment of the pollen-stigma interaction is proposed to involve a basal compatibility system that enables compatible pollen to be recognised by the receptive stigma. Divergence of components involved in this system could facilitate the establishment of prezygotic breeding barriers that would limit wasted mating opportunities, restrict interspecies gene flow and contribute to reproductive isolation. A diverse family of small secreted cysteine-rich proteins (CRPs) found in the pollen coat of members of the Brassicaceae, the pollen coat proteins (PCPs), are emerging as important regulators of the pollen-stigma interaction. One class of PCPs isolated from the pollen coat of *Brassica oleracea*, the PCP-Bs, have previously been described, but their function was unknown. In this study, four putative *Arabidopsis thaliana* PCP-B-encoding genes were identified, determined to be gametophytically expressed during the late stages of pollen development and confirmed as pollen coat proteins. Bioassays utilising single and multiple *pcp-b* gene knockouts revealed that *At*PCP-Bs function in the early stages of post-pollination. To identify the stigmatic targets of AtPCP-B ligands, a series of protein-protein interaction (PPI) assays were carried out with heterologously expressed AtPCP-Bs and isolated stigmatic proteins. To provide insight into the evolutionary characters of PCP-Bs, phylogenetic analysis and molecular evolutionary study revealed evidence of positive selection acting on sites of genes encoding PCP-Bs and PCP-B-like proteins. Such evidence suggests that *At*PCP-Bs are important components of the basal compatibility system by establishing a molecular dialogue between compatible pollen grains and the stigma. Proteomic analyses of pollen coat from *Arabidopsis thaliana* and *Brassica oleracea* uncovered large numbers of small CRPs, which may act as important regulatory factors of pollen-stigma interaction. This project shed new light on the biological and evolutionary significance of pollen coat CRPs in plant reproductive signalling.

Contents

Chapter 1 Introduction

The process of sexual reproduction and many of its associated features in higher plants have played a central role in the diversification of angiosperms. Prezygotic reproductive barriers, built on the processes of post-pollination events, significantly limit gene flow between populations and eventually lead to speciation. After the transfer of pollen to the female reproductive organ, post-pollination events are stringently regulated by a series of signalling proteins. Thus the evolution of genes encoding these regulators has contributed to the success of angiosperms. In this chapter, discussion of the topic is mostly restricted to the members of Brassicaceae that contains important crop plants and model species. In this family, the stigma is highly selective with respect to the pollen landing on its dry surface where only interspecies or intraspecies compatible pollen can be accepted. During the last several decades, there has been accumulating evidence that small cysteine-rich proteins (CRPs) act as important regulators during multiple stages of plant sexual reproductive. This project focuses on a class of pollen coat CRPs, the pollen coat protein B class (PCP-Bs), where the function of these genes and their evolution will be studied. The search for potential stigmatic targets of PCP-Bs and characterization of other pollen coat protein components was also a goal of this project.

1.1 The evolution of angiosperms

The sudden origin and rapid diversification of angiosperms was once described by Charles Darwin as an ‘abominable mystery’ and ‘perplexing phenomenon’ (Darwin & Seward, 1903). Evidence from the fossil record suggests that the origin of angiosperms dates to the early Cretaceous (130-136 million years ago) (reviewed in Friis *et al.*, 2006), and fossils of angiosperm-like pollen have recently been discovered from the Middle Triassic (Anisian, 247.2-242.0 million years ago) (Hochuli & Feist-Burkhardt, 2013). In 2013, a draft genome of *Amborella trichopoda* was published, which revealed that *Amborella* is the only living sister taxon of all extant flowering plants (Albert *et al.*, 2013). Intra-genomic, syntenic and comprehensive phylogenomic analyses of *Amborella trichopoda* revealed that shortly before the diversification of angiosperms, an ancient whole genome duplication (WGD) event occurred, which led to the innovation of novel genes and likely played a

central role in the dominance of angiosperm lineages on our planet (Jiao *et al.*, 2011; Albert *et al.*, 2013). Following the origin of angiosperms, a rapid diversification and ecological dominance lead to more than 350,000 species that now occupy most habitats on the earth today. By utilising developmental biology and genetics, phylogenetics and genomics, as well as palaeobotany, scientists have revealed many aspects of the mysterious history of angiosperms evolution. Fundamentally, the study of angiosperm sexual reproduction supports the understanding of the molecular and evolutionary basis of angiosperm diversification.

1.1.1 The evolution of sexual diversity in angiosperms

As the most important evolutionary innovation of angiosperms, flowers are the reproductive structure and vary more than equivalent organs in other organisms. To achieve mating success, the sexual reproduction system evolved diverse mating strategies to ensure a high degree of fitness in different ecological niches. There are two major transitions during the evolution of the sexual systems in angiosperms: the transition from hermaphroditism to dioecy and the transition from cross-fertilisation to self-fertilisation (Barrett, 2002). As commonly accepted, inbreeding, especially self-fertilisation can lead to the accumulation of harmful mutations resulting in inbreeding depression (Stebbins, 1957), which is considered as the major selective force driving plant mating strategies (Charlesworth & Charlesworth, 1987). Consequently, sexual polymorphisms have evolved to impede self-fertilisation. In the most predominant flower type, hermaphroditic flowers, the reproductive organs of two genders can be separated spatially (herkogamy) and temporally (dichogamy) to reduce the chance of self-pollination, such as heterostyly, enantiostyly and flexistyly. In addition to structural polymorphisms, outcrossing of hermaphroditic plants can also be enforced by self-incompatibility (SI) (to be discussed in detail in 1.3.3.1), a molecular mechanism that enables an individual plant to recognise and reject its own pollen. The evolution of sexual dimorphism, especially dioecy, is puzzling because plants are immobile organisms and the fertilisation could be severely impacted by a reduction in the availability of pollination vectors. The most commonly accepted model for the selective mechanism operating in this case is inbreeding avoidance that prevents self-fertilisation. For example, in diclinous flowers, their strictly unisexual (dioecy) or

partially unisexual (gynodioecy and androdioecy) morphology builds a natural spatial barrier between the two genders thus preventing self-fertilisation (Barrett, 2002). However, though dioecy is widely distributed in half of the angiosperm families, only about six percent of angiosperm species are dioecious (Renner & Ricklefs, 1995). Theoretical studies have indicated that the primary factor that governs the evolution of dioecy is the rate of self-fertilisation and inbreeding depression in hermaphrodite populations (Charlesworth, 1999). Biological factors that are responsible for increasing the incidence of self-fertilisation are not fully understood but ecological factors such as harsh and stressful environmental conditions have been considered to affect the evolution of separated sexes (reviewed in Barrett, 2010).

The transition from predominant outcrossing to a high level of selfing is one of the most frequent evolutionary events in plants (Barrett, 2002). The shifting of mating strategy to selfing involves a series of morphological and functional changes of reproductive organs (selfing syndrome), which leads to a reduction of gene flow between individuals, followed by differentiation between populations and accumulation of reproductive isolation factors (Wright *et al.*, 2013). Self-fertilisation has been considered as an evolutionary dead-end because of the associated effects of inbreeding depression, however, natural selection does not predict the future but only chooses the immediate fitness. Thus when outcrossing benefits long-term fitness of plants (Wright & Barrett, 2010), self-fertilisation benefits from its short-term evolutionary advantages. Self-fertilisation not only releases plants from the requirement of mating partners but also enables plants to rapidly colonise unoccupied space. Although accumulating evidence suggests that selfing may facilitate speciation (Wendt *et al.*, 2002), however, in fact, only 10-15% of species are predominantly self-fertilising, whilst around 42% of species studied present as mixed mating (outcrossing rate falls between 20-80%) (Goodwillie *et al.*, 2005). There are two hypotheses explaining the intriguing reasons behind the evolution of self-fertilisation: 'reproductive assurance' suggests that selfing guarantees pollination when the availability of pollinators or mating partners are scarce or unstable (Darwin, 1876), while 'automatic selection' suggests that the selfing populations hold the genetic transmission advantage by acting as both paternal and maternal parents of the next generation (Fisher, 1941). Studies on mating systems across a range of species reveal that those having biotic pollination present a continuous variation in outcrossing rate, while the species with abiotic pollination are mostly predominantly selfing or outcrossing (Figure 1.1) (Vogler & Kalisz, 2001; Barrett, 2002;

Goodwillie *et al.*, 2005). This distribution supports the 'reproductive assurance' hypothesis, which proposed that the presence of pollinators or mating partners has relatively less impact on the reproduction in species with abiotic pollination. On the species level of evolution, self-incompatibility (SI) is a trait that benefits long-term fitness, enhancing the diversification of a species. In contrast, self-fertilisation acts as a trait that guarantees the successful pollination in the short term, which however can lead the species to extinction over the longer term (Goldberg *et al.*, 2010). This species level of evolution is clearly evident in the Solanaceae where the diversification rate in SI species is significantly higher than selfing species, whilst extinction occurs more frequently in selfing species than SI species (Goldberg *et al.*, 2010). However, the high prevalence of mixed mating species in some of angiosperm families increases the difficulty of reproductive compatibility studies.

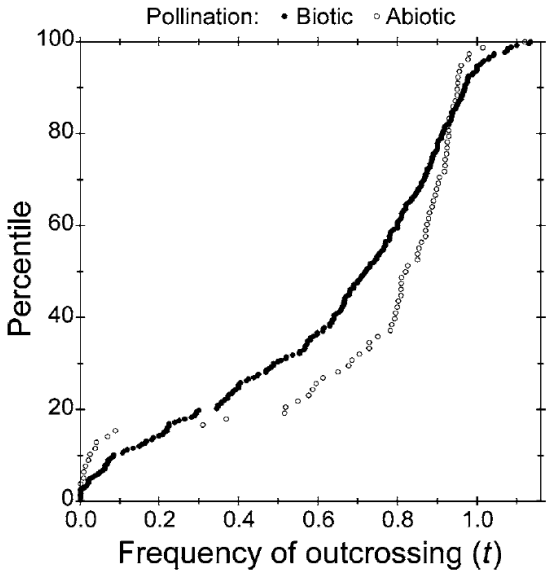


Figure 1.1 | Comparison of estimated outcrossing ratio (*t*) in biotic and abiotic pollinating species. The graph presents the percentile to the estimated *t* for 267 biotic (closed circles) and 78 abiotic pollinating species (open circles). For each pollination mode, species were ranked (percentiles) based on estimated *t*. The two distributions show the functional link of mating systems of species and their pollination biology. Figure adapted from Goodwillie *et al.*, 2005.

1.1.2 The evolution of reproductive proteins in plants

Reproductive proteins are encoded by the genes that mediate sexual reproduction. In plants, reproductive proteins are broadly defined as those which act after pollen attachment on the female reproductive tissue and have roles in regulating pollination (events following pollen attachment), fertilisation and seed development. Consequently, reproductive protein evolution influences species fitness. The reproductive protein-encoding genes are more

divergent than non-reproductive related genes (Vacquier, 1998). The surprisingly high level of diversity and divergence indicates that reproductive proteins may evolve as the consequence of adaptive evolution. A series of statistical tests have been developed to uncover some common molecular features of this adaptive evolution process (Swanson & Vacquier, 2002). The most widely used class of tests defines the number of non-synonymous substitutions per possible non-synonymous codon sites as d_N, whereas the number of synonymous substitution per possible synonymous codon sites as d_S. The ratio of d_N and d_S ($\omega = d_N/d_S$) indicates the selection type. $d_N/d_S=1$ represents a neutral amino acid changing substitution, whereas $d_N/d_S>1$ demonstrates positive selection for amino acid change (reviewed in Yang & Bielawski, 2000). There are a few reported cases of reproductive proteins in angiosperms known to be under positive selection (reviewed in Clark *et al.*, 2006). In some species, the early pollen-stigma interaction is regulated by a self-recognition system termed as self-incompatibility (SI) (see section 1.2.2). In the Rosaceae and Solanaceae SI is gametophytically determined (GSI), which involves a stylar-expressed *S*-locus protein with RNase activity (*S*-RNase) (Anderson *et al.*, 1986; McClure *et al.*, 1989) and the pollen component an F-box protein closely linking to *S*-RNase (SLF) (Lai *et al.*, 2002; Sijacic *et al.*, 2004). This GSI system is regulated by the interaction of *S*-RNase and SLF variable domains, which triggers downstream signalling that impedes pollen tube growth (McClure, 2004). *S*-RNase and SLF are highly polymorphic (Entani *et al.*, 2003; Kato & Mukai, 2004) and the statistical model $d_N/d_S>1$ supports that positive selection acts on the adaptive diversification of their functional domains (Takebayashi *et al.*, 2003; Ikeda *et al.*, 2004). In the Brassicaceae, SI is sporophytically determined (SSI), and is primarily controlled by two genes on the same *S*-locus, the *S*-locus cysteine-rich protein (SCR) and *S*-locus receptor kinase (SRK) (reviewed in Iwano & Takayama, 2012) (the functions of these proteins in SI will be reviewed in section 1.3.3.1). Similarly, amongst populations, these proteins are highly polymorphic (Takebayashi *et al.*, 2003). Based on the statistical model $d_N/d_S>1$, the adaptive diversification of SCR and SRK (and also the SRK related protein, SLG) is also driven by positive selection (Sato *et al.*, 2002; Takebayashi *et al.*, 2003; Guo *et al.*, 2011). Additionally, a proteomic analysis of *Arabidopsis thaliana* pollen coat revealed a gene cluster containing six lipid binding oleosin-like genes (Mayfield *et al.*, 2001). Individual oleosin-like proteins show high polymorphism amongst ecotypes and a syntenic region in *Brassica oleracea*, which implied that they are involved in adaptive evolution or even speciation (Mayfield *et al.*, 2001).

Though the SI system provides evidence for positive selection of reproductive proteins, there is a fundamental question that has not been answered by the current evidence: what is the molecular mechanism of the evolution of gamete recognition proteins in angiosperms? A substantial body of research has focused on the evolution of gamete recognition proteins in animals (Vacquier, 1998; Swanson & Vacquier, 2002). One of the paradigm cases in this field is the abalone fertilisation system that has evolved strong barriers to cross-fertilisation with its closely related species (reviewed in Lessios, 2011). This well-characterised gamete recognition system in abalone involves the interaction of sperm lysin and its receptor, vitelline envelope receptor for lysin (VERL), a large glycoprotein with 22 tandem repeats (Lewis *et al.*, 1982; Swanson & Vacquier, 1997). The amino acid sequences of lysins from different species of abalone were found to be highly divergent and rapidly evolved under positive selection (Metz *et al.*, 1998). In contrast, only two repeats in VERL evolve by positive selection but the other 20 repeats are not subject to positive selection but evolve neutrally instead (Swanson & Vacquier, 1998; Galindo *et al.*, 2002; Galindo *et al.*, 2003). Based on the characterisation of the molecular evolution in this species-specific fertilisation system, a model explaining the mechanism of speciation was hypothesised. After a population split, mutation of one of the 22 VERL repeats results in a lower affinity for lysin binding, but fertilisation still occurs due to the presence of 21 unchanged repeats. The redundancy of the VERL repeating domains diminishes the consequence of the mutation on fitness and enables gene conversion that spreads the mutant repeat through the VERL gene and eventually leads to concerted evolution (Elder & Turner, 1995; Metz & Palumbi, 1996; Swanson & Vacquier, 1998). Thus, the molecular mechanism of the adaptive evolution of lysin could be explained as a continuous selective force from the ever-changing VERL (Figure 1.2) (Swanson & Vacquier, 2002). This hypothesis represents the only model that utilised sequencing data to explain the mechanism of coevolution between gamete recognition protein pairs. In plants, there is no similar study where analysis of sequence data has been interpreted to support this hypothesis, however several protein pairs have been demonstrated to be coevolving and involved in male-female recognition events (Sato *et al.*, 2002; Guo *et al.*, 2011).

The potential forces driving positive selection can be broadly divided into two categories, endogenous forces of the species reproductive system and the exogenous forces from the loci involved in pathogen resistance. One example of positive selection driven by

endogenous forces involves the SI system, which is thought to be driven by the avoidance of inbreeding. Once inbreeding depression is strong enough to establish an SI system, the evolution of gene loci corresponds to negative frequency-dependant selection (Wright, 1939). Under this selection, pollen with rare alleles are more likely to be accepted by the female reproductive tissue, which facilitates the diversification of alleles within populations and results in high polymorphism of SI-related genetic loci. For example, 49 and 44 *S* haplotypes (see section 1.2.2) have been reported in *Brassica olearacea* and *Brassica rapa* (Oikawa, *et al.*, 2011). The highly levels of polymorphism at loci controlling SI then further facilitates inbreeding avoidance (Clark *et al.*, 2006). Pathogen resistance, in contrast, is hypothesised to be a potential exogenous force. Constant pathogen attack may drive gamete surface proteins to change to defence, whereas the microbial proteins need to constantly evolve to recognise the new host surface (Vacquier *et al.*, 1997).

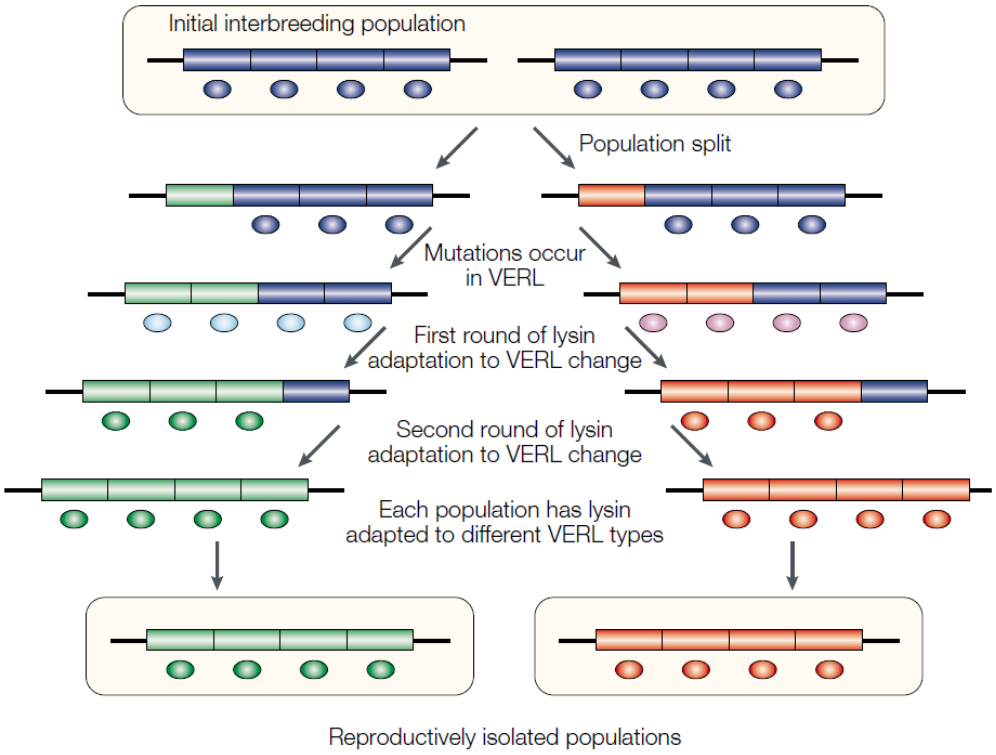


Figure 1.2 | Abalone Lysin-VERL coevolution model. VERL and lysin is represented as coloured bar and coloured circles respectively. Figure adapted from Swanson & Vacquier, 2002.

1.1.3 Reproductive isolation in plant speciation

To further explain the driving force of angiosperm diversification, it is necessary to understand the process of plant speciation. Although the definition of what constitutes a plant species has been challenged due to the common occurrence of interspecific hybrids and the continuous variations in phenotype between plant groups (Mishler & Donoghue, 1982; Arnold *et al.*, 1999), recent work has revealed that most of the taxonomic species and phenotypic clusters correspond to the reproductively isolated lineages (controlled by post-pollination events), which indicates that the vast majority of plant species do represent reproductively independent groups (Rieseberg *et al.*, 2006). For plants that mainly reproduce by sex, the formation of species is achieved by the evolution of reproductive isolation that impedes gene flow between formerly interbreeding populations. Reproductive barriers facilitate the accumulation of genetic variation by reducing gene flow between populations, which sharpens their boundaries, promotes adaptive traits and eventually leads to complete reproductive isolation (Rieseberg & Willis, 2007). Most plant species are formed and maintained by multiple reproductive barriers, including pre-pollination barriers, post-pollination prezygotic barriers and postzygotic barriers. Pre-pollination barriers reduce the possibility of the pollen from one species being transferred to another species. These barriers are built up by spatial, temporal, morphological, mechanical and pollinator isolation. Prezygotic barriers act before fertilisation, resulting in the privileged acceptance of conspecific pollen. Postzygotic barriers act after fertilisation, resulting in hybrid inviability and hybrid breakdown that reduces or fails reproduction in the next generation. Several studies have provided evidence that early-acting reproductive barriers, such as ecographic and pollinator isolation (Ramsey *et al.*, 2003; Husband & Sabara, 2004; Kay, 2006), as well as prezygotic barriers, such as mating system isolation (Martin & Willis, 2007), contribute more than the postzygotic barriers to isolation. Though the barriers of geographic isolation has been thought to contribute most to total isolation, prezygotic barriers are emerging as very important factors in plant speciation and can be possibly manipulated based on the molecular and cellular regulatory mechanisms that have been studied intensively. Studies of rapidly evolving regions of reproductive protein-encoding genes and the coevolution of interacting molecules may shed new light on the molecular basis of interspecific incompatibility and the early stages of intraspecies gamete recognition. However, the large number of prezygotic barriers and their potential

interactions involved in different steps of pollination make it a huge challenge to identify the molecules and their roles in reproductive isolation.

1.1.4 Evolution by gene duplication

In addition to the diversification of molecules involved in reproductive barriers, a large proportion of plant speciation is triggered by polyploidy. Gene duplication is the direct consequence of polyploidy and acts as an important factor in the diversification of angiosperms by providing new raw genetic material. Evolution after gene duplication contributes to the emergence of new gene functions that facilitate the generation of novel morphologies and physiologies. In addition to the ancient whole genome duplication (WGD) events in seed plants and angiosperms, the more recent WGDs in the *Arabidopsis* lineage resulted in the large proportion of duplicated genes in *Arabidopsis thaliana* (Blanc *et al.*, 2000; Jiao *et al.*, 2011; Albert *et al.*, 2013). Gene duplication may occur through three major mechanisms: segmental duplication, tandem duplication and retroposition (reviewed in Zhang, 2003). Segmental duplications have been revealed more frequently in plants as a result of WGD events that produced duplicated blocks on the chromosomes within their genomes (Zhang, 2003; Cannon *et al.*, 2004). Although the duplicate blocks appear to be highly degenerated and disrupted, they can be recognised and reveal the polyploidy history of the diploid genomes (Blanc *et al.*, 2000; Simillion *et al.*, 2002; Blanc *et al.*, 2003). Tandem duplication is usually caused by unequal crossing-over. The recent origin and gene conversion result in the high similarity among the tandemly arrayed gene copies. For both segmental and tandem duplications, if introns are present in the original genes they are generally retained in the duplicated copies. In contrast, introns are absent if gene duplication is the result of retroposition, a consequence of reverse-transcription of mRNA to a cDNA that then becomes inserted into genome (Zhang, 2003). Thus, retroposition events can be recognised by their lack of introns and regulatory regions, as well as the presence of poly A tracts and flanking short direct repeats (Kaessmann *et al.*, 2009). Another character of retroposition is that the duplicated gene is usually not linked to the original gene due to the random insertion of reverse-transcripted cDNA into the genome.

After the duplication occurs in an individual, the duplicated gene can be fixed or lost in the population. As a consequence, the birth and death of genes are commonly observed during

genome evolution. Pseudogenisation is the process by which a functional gene becomes a pseudogene, which occurs due to the functional redundancy generated by gene duplication. Within the first several million years after duplication, if the disruption of structure and function in the duplicated gene is not under any selection, it gradually losses its function or becomes unexpressed. After a long time, the pseudogene will be deleted or become extremely divergent. Only the more recently pseudogenised products can be recognised. A pseudogene has been traditionally defined as a sequence of genomic DNA that displays degenerative features such as premature stop codons and shifted reading frames that cause defects in expression (Li *et al.*, 1981). However, accumulating cases of functional putative pseudogenes have been reported (Balakirev & Ayala, 2003). Instead of undergoing pseudogenisation, the duplicated genes can be maintained in the genome. The function of the fixed duplicated gene will determine its long-term evolutionary fate. In theory, there are three possible outcomes during the evolution of the duplicated gene copies: 1. Gene conservation, the mutation is deleterious thus the gene sequence and function was maintained by purifying selection. 2. Subfunctionalisation, the functional capacity of both copies was reduced and became partially compromised to the level of their ancestral gene. 3. Neofunctionalisation, the original function of one copy was maintained while the other copy acquired beneficial novel function that could be preserved by purifying selection (Lynch & Conery, 2000; Zhang, 2003). In fact, there has been a series of models proposed on the maintenance and evolution of gene duplications (Innan & Kondrashov, 2010). As presented in a review by Innan and Kondrashov (2010), the evolutionary process after gene duplication was divided into three phases: the fixation phase, fate-determination phase and preservation phase (Figure 1.3). The models of evolution after gene duplications were classified into four categories based on their selective forces on the new gene copy during the fixation phase. However, due to the significant overlap between these models, it is almost impossible to fit any specific duplicated gene group into an individual hypothesised model. Nevertheless, we are still sure that to gain any functional or structural divergence, it is essential for the duplicates to undergo diversifying selection at some stages during their post-duplication evolutionary history.

The rapid evolution of reproductive proteins is a hallmark of species-specific recognition systems (reviewed in Swanson & Vacquier, 2002). Thus, the studies on duplication and evolutionary history of genes involved in plant sexual reproduction may provide valuable clues for understanding the origin and evolution process of reproductive barriers. When

combined with gene functional analyses, these studies could further inspire the discovery of the molecular basis of plant reproductive regulation and speciation.

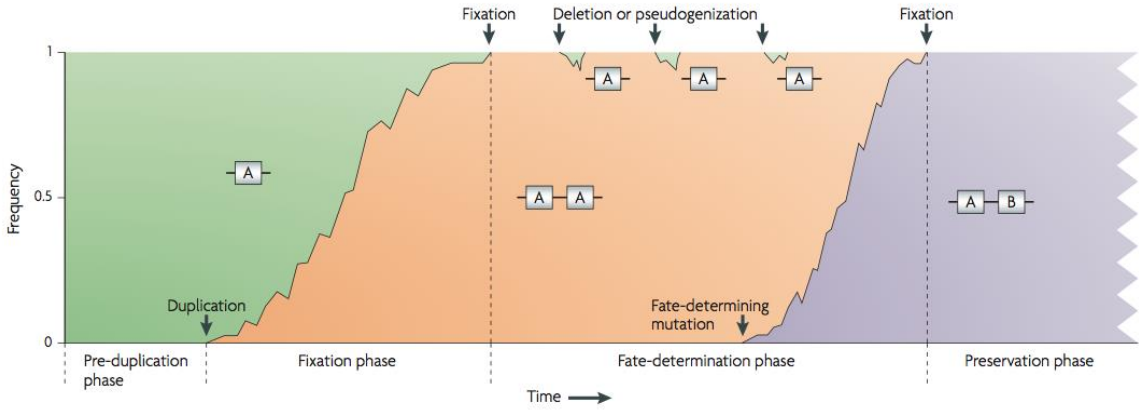


Figure 1.3 | Phases from shortly before the duplication event occurs to the preservation of a duplicated gene. 'A' represents the single-copy genotype (Innan & Kondrashov, 2010).

1.2 Early stages of pollination

Sexual reproduction in angiosperms results from the fertilisation of female gametes in the ovule by male gametes delivered by pollen grains. The process of sexual reproduction in angiosperms is characterised by 'double fertilisation' which involves fusion of two male and two female gametes (Weterings & Russell, 2004). Successful fertilisation is enabled by the pollination process that transports male gametes via a pollen tube through the female reproductive tissues to the ovule where the female gametes reside. The process of sexual reproduction in angiosperms leading up to fertilisation can be divided into four steps: pollen adhesion, pollen hydration, pollen tube growth and pollen tube attraction followed by sperm delivery. After entering the ovule, the sperms migrate towards the two female gametes, which is followed by the cellular and genetic fusion of male and female gametes (karyogamy), and re-initiation of the new zygotic cell cycle (Figure 1.4). One sperm fused with the egg cell to form the zygotes and the second sperm fuses with the central cell to produce endosperm and subsequent development of this tissue provides the nutrients needed for growth of the diploid zygote (reviewed in Friedman, 1998). Sexual reproduction in angiosperms is strictly regulated by intercellular molecular communication between male and female reproduction structures (reviewed in Dresselhaus & Franklin-Tong, 2013), which contribute to the reproductive barriers that impede interspecies mating.

In some species, this molecular regulation can be targeted by components of the self-incompatibility system as a mechanism to avoid self-fertilisation.

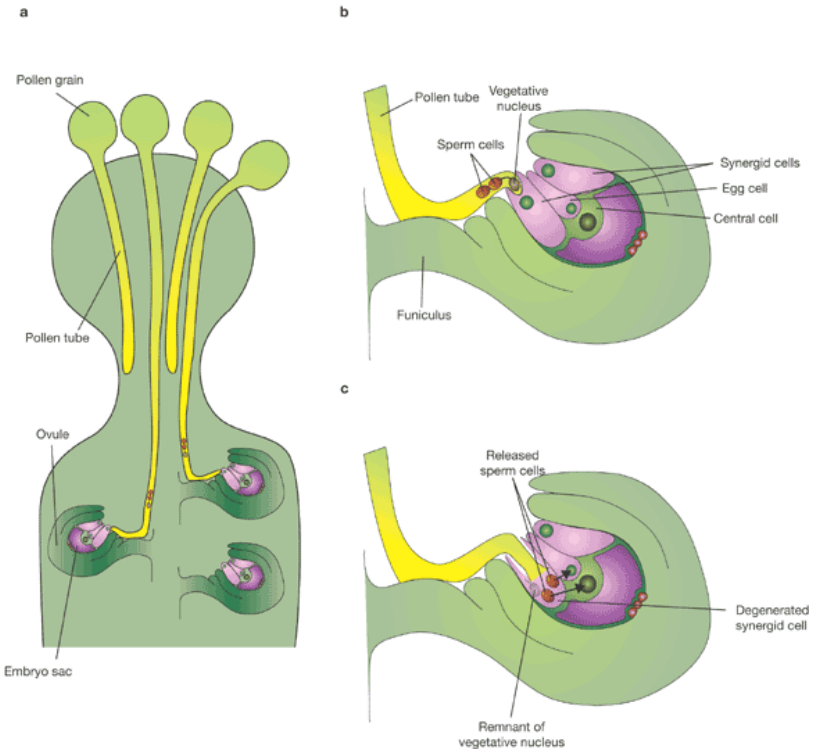


Figure 1.4 | Double fertilisation in angiosperms (Twell, 2006). (a) Diagram of pollen germination and elongation into the pistil. (b) Each pollen tube contains two sperms cells and a vegetative nucleus. The pollen tube is attracted towards the micropyle - the opening of the ovule. (c) Two sperm cells were discharged from the pollen tube into one of the synergid cells of the embryo sac. This process is followed by double fertilisation: one sperm cell fuses with the egg cell to produce embryo and the other with the central cell to produce endosperm of the seed.

1.2.1 Pollen-stigma interaction in Brassicaceae

Pollination initiates when the pollen grain is captured by the stigma (Figure 1.5). Amongst the member of Brassicaceae, the initial steps of pollen capture and adhesion rely on the pollen outer wall, the exine (Gaude & Dumas, 1984), a highly sculptured structure mainly composed of sporopollenin (reviewed in Piffanelli *et al*., 1998). Formation of exine is essential for the adhesion of the pollen grain to the stigmatic surface (Nishikawa *et al*., 2005; Dobritsa *et al*., 2009). In addition to the exine, the pollen coat (also known as tryphine, also contributes to the pollen adhesion (Heslop-Harrison, 1968) (see section 1.4). The stigma is the receptive terminal of a carpel or a group of carpels in the gynoecium, connecting to the ovary by the style. Though the size and structure varies among species, stigmas can be broadly classified into two groups: wet stigmas and dry stigmas (Heslop-

Harrison & Shivanna, 1977). Wet stigmas are covered by a viscous secretion that contains polysaccharides, proteins, lipids and pigments (e.g. *Nicotiana* and *Petunia* in Solanaceae). Dry stigmas lack a surface exudate and possess a layer of intact papillar cells that are covered by a waxy cuticle and proteinaceous pellicle (e.g. Brassicaceae) (Edlund *et al.*, 2004). The stigmas of *Senecio squalidus* on the other hand possesses a semi-dry surface that is covered by a cuticle that not continuous to the base of papillar cells, where small amount of exudate is present when the stigma is mature (Hiscock *et al.*, 2002). Though this semi-dry stigma is commonly observed in the Asteraceae, its molecular and evolutionary relationship with common dry and wet stigmas is unclear.

Once captured and adhered, the pollen undergoes hydration on the stigma surface. Though the hydration process varies in different families with different stigma types, here, in this project, the topic focuses on members of the Brassicaceae family which have stigmas of the dry type. To hydrate the desiccated pollen grain on the surface of dry stigma, the pollen coat mobilises towards the stigma surface to form a 'foot' at the contact site (Figure 1.5). It is believed that the pollen coat (see section 1.4 for detailed consideration) and stigmatic cuticle layer form a hydraulic conduit for water transport from the stigma to pollen grain (Elleman *et al.*, 1992; Dickinson, 1995; Edlund *et al.*, 2004). Dry stigmas provide a highly discriminatory environment that reduces the probability of capturing pathogenic spores and dust compared to wet stigmas and also reduces stigma clogging and energy wastage from the hydration and potential pollen tube growth of heterospecific or incompatible pollen. The evolution of the dry stigma makes pollen hydration a crucial checkpoint in the reproductive process, when a 'decision' needs to be made by a molecular dialogue between the pollen grain and the stigmatic papillar cell in order to accept the compatible pollen grain. In the Brassicaceae, the hydrated pollen produces a pollen tube that penetrates the stigma cuticle and elongates between the two layers of cell wall of the finger-like papillar cell (Elleman *et al.*, 1992). At the base of the papillar tissue, the pollen tube penetrates the style and grows towards the ovule (Figure 1.5). Guidance of the pollen tube growth includes correct directional growth through sporophytic tissues of the style and appropriate targeting to permit the tube to enter the ovule, this final step being mediated by signalling from the synergid cells (Higashiyama *et al.*, 2001; Okuda *et al.*, 2009; Takeuchi & Higashiyama, 2012). Once the pollen tube gains access to the ovule, the pollen tube releases the two sperm cells to complete the double fertilisation. The entire pollination process is strictly regulated by the signalling factors produced by the male and female

reproduction tissue (Liu *et al.*, 2013; Wang T *et al.*, 2016). These signalling molecules are the basis of reproductive barriers that facilitate plant speciation and self-incompatibility in some species.

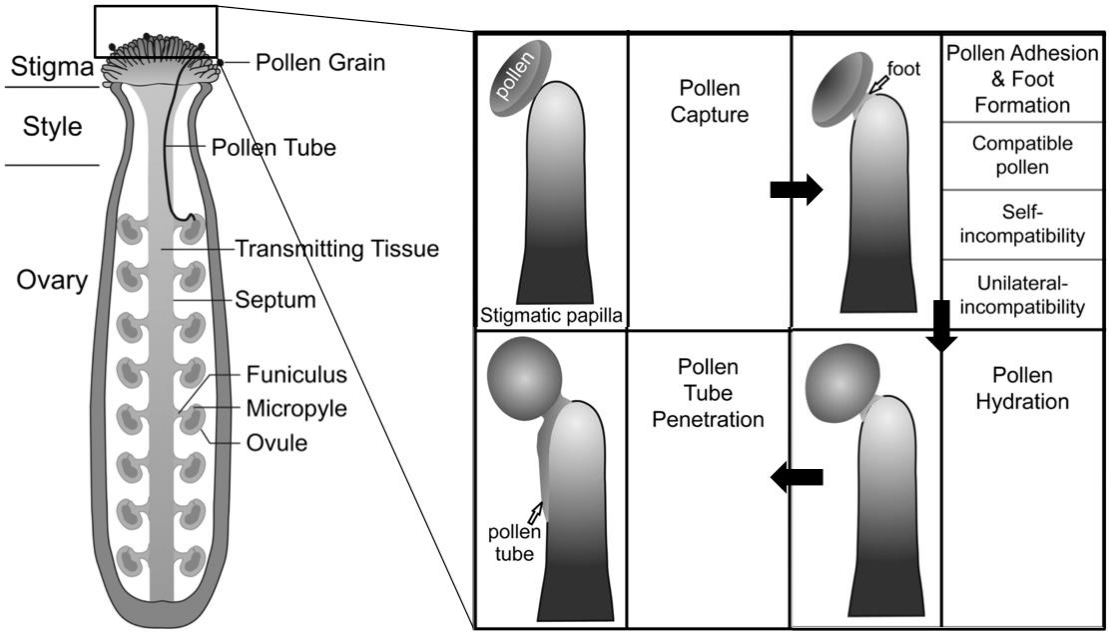


Figure 1.5 | Schematic of an *Arabidopsis thaliana* pollinated pistil and diagrams illustrating the four initial stages of post-pollination. Black boxes indicate the initial stages of pollen-stigma interaction. Figure adapted from Chapman & Goring, 2010.

1.2.2 The pollen-stigma self-incompatibility and basal compatibility during the early stages of pollination in the Brassicaceae

A pollen grain in contact with the stigma can be either compatible or incompatible at the interspecific or intraspecific level. The pollination process described in the previous section only completes when the adhered pollen grains are compatible with the stigma. At the interspecies or inter-population level, multiple highly selective prezygotic barriers act in heterospecific reproductive tissue. Accumulation of genetic differences sharpens the boundaries between population groups or closely related species, which eventually leads to complete reproductive isolation. On the intraspecies level, many angiosperm species have evolved a self-incompatibility (SI) system, a well-studied molecular genetic strategy for restricting self-fertilisation. In gametophytic self-incompatibility (GSI) system existing in Solanaceae and Papaveraceae, the SI phenotype of the pollen is determined by its own

gametophytic haplotype. One of the mechanisms of GSI is regulated by the female determinant ribonuclease termed as *S*-RNase and the male determinant 'F-box' family (reviewed in Kao & Tsukamoto, 2004). The incompatible pollen tube elongation has been observed to be casued by the interaction of identical female and male *S* alleles, which may trigger the ribosomal RNA (rRNA) in the growing pollen tube (Sijacic *et al*., 2004). Another mechanism of GSI has been well studied in *Papaver rhoeas*, in which the pollen *S*-determinant (PrpS) interacts with the stigma *S*-determinant (PrsS) secreted to the pistil surface (Wheeler *et al*., 2009). The interaction of PrsS and PrpS triggers a signalling cascade that increases intracellular Ca^{2+} concentration, resulting in the inhibition of pollen tube growth and eventually programmed cell death (PCD) (reviewed in Eaves *et al*., 2014). In sporophytic self-incompatibility (SSI) system that is well studied in Brassicaceae, the SI phenotype of pollen is determined by the sporophytic diploid genotype of anther (Schopfer *et al.*, 1999; Shiba *et al.*, 2001). The male determinant, *S*-locus cysteine-rich protein (SCR), is derived from the anther tapetum. The female determinant of SSI is a transmembrane protein termed *S*-locus receptor kinase (SRK), possessing a diverse extracellular domain and an conserved intracellular kinase domain (Takasaki *et al*., 2000). The ligand-receptor interaction between SCR and SRK results in the phosphorylation of the intracellular domain of SRK, which triggers downstream signalling pathway mediated by an arm repeat-containing protein (ARC1) with E3 ubiquitin-ligase activity (Stone *et al*., 2003). The *S*-locus gene encoding SRK is linked to another *S*-locus gene *SLG* that encodes *S*-locus glycoprotein (SLG) (Takasaki *et al*., 2000). SLG possesses a highly similar sequence with SRK protein and may enhance the SI response by acting as the co-receptor with SRK to SCR (Takasaki *et al*., 2000). In *Brassica rapa*, an *M*-locus protein kinase (MLPK) was found to bind to SRK and form an SRK-MLPK receptor complex to enhance SI response (Kakita *et al*., 2007). The activation of ARC1 was discovered to facilitate the self-pollen rejection by negatively regulating the exocyst subunit Exo70A1, which inhibits the anchoring of stigmatic vesicles to the membrane at the pollen contact site (Samuel *et al*., 2009).

In the Brassicaceae family, there are both SI and self-fertilising species. Typical examples of SI species include *Brassica oleracea* and *Arabidopsis lyrata*, whereas the self-fertilising model plant *Arabidopsis thaliana* has lost the SI system possessed by its ancestors (Tsuchimatsu *et al.*, 2010). During the last several decades, a large amount of research has been published describing the mechanism of self-incompatibility (SI), yet fewer studies

have addressed the molecular factors involved in this very initial stage of pollen-stigma recognition. In plants species with dry stigmas such as members of Brassicaceae, the pollen hydration process on the stigma can act as the very front line of the prezygotic barrier. By impeding heterospecific pollen hydration the plant is able to reduce energy wastage and maintain reproductive tissue fitness. By being located at the interaction frontier, the pollen coat and stigmatic papillar cells evolved to communicate via a molecular dialogue that enables the stigma to recognise and select compatible pollen. Mutational studies in *Arabidopsis thaliana* have demonstrated that several pollen-borne factors are influencing the efficiency of pollen hydration. For instance, the elimination of very long chain lipids from the pollen coat in *eceriferum* (*cer*) mutants resulted in the failure of pollen hydration (Preuss *et al.*, 1993; Hulskamp *et al.*, 1995; Fiebig *et al.*, 2000). Instead of playing any specific role in the molecular communication, these lipids are more likely essential for the solubilisation and mobilisation of other pollen coat components (Hulskamp *et al.*, 1995) or for the establishment of hydration conduits (Elleman *et al.*, 1992; Dickinson, 1995; Wolters-Arts *et al.*, 1998; Edlund *et al.*, 2004). It has also been revealed that pollen hydration is impaired by the absence of extracellular lipase 4 (EXL4) and an oleosin-domain-containing glycine-rich protein (GRP17) in pollen coat, which may cooperatively facilitate water transport from papillar cells to pollen grains by altering the lipid composition at the pollen-stigma interface (Mayfield & Preuss, 2000; Updegraff *et al.*, 2009). Studies of SI in *Brassica* identified several groups of small pollen coat cysteine-rich proteins (CRPs) (Doughty *et al.*, 1993; Hiscock *et al.*, 1995; Doughty *et al.*, 1998; Schopfer *et al.*, 1999; Doughty *et al.*, 2000; Takayama *et al.*, 2000a; Shiba *et al.*, 2001). Rather than stabilising the biophysical properties of pollen coat, these proteins act as ligands and bind to a series of stigmatic proteins (see sections 1.3.3 and 1.4). This model suggests that the early stages of self-recognition are likely to be mediated by signalling molecules produced by reproductive tissue.

1.2.3 The cellular and molecular responses of the stigma during pollen acceptance and rejection

To gain an understanding of the molecular basis of basal compatibility in the pollen-stigma interaction, consideration of current knowledge of the cellular and molecular events during

the early stages of pollination may provide us with some clues. During the initial stages of pollination and SI events, the cellular and molecular responses to the pollen grain occur close to the stigmatic papillar cell surface near the pollen grain attachment region. In species with dry stigmas, the stigma surface cells need to undergo dramatic changes to enable the output of water and allow the pollen tube to penetrate the cell wall. In *Arabidopsis* and *Brassica*, after the adhesion of compatible pollen, the outer element of the papillary cell wall expands around the 'foot' of the pollen coat, which enables the pollen tube to grow into the space between the inner and outer element of the wall (Clarke *et al.*, 1979; Elleman *et al.*, 1992). In *Papaver rhoeas*, beneath the attached compatible and incompatible pollen grains, the stigmatic surface secrets electron-lucent material under the cuticle, causing the cuticle to become detached from stigmatic cell wall (Elleman & Dickinson, 1996). In *Arabidopsis* and *Brassica*, close examination of the stigmatic plasma membrane and cell wall under compatible pollen grains demonstrated the secretion of vesicles / multivesicular bodies that move from the cytoplasm to the apoplast (Elleman & Dickinson, 1996; Safavian & Goring, 2013). Penetration of the pollen tube through the stigmatic cuticle layer and cell wall has been proposed to require enzymatic modification which allows pollen tube growth. The discovery of significant esterase activity in *Brassica* pollen and in stigmas added support to the suggestion that the esterases, including cutinases, are likely to play essential roles in this process. Indeed, blocking of pollen tube penetration and growth occurred after the treatment of stigmatic cells with a serine esterase inhibitor (Hiscock *et al.*, 1994, 2002; Lavithis & Bhalla, 1995).

A recent study suggested that the dynamics of the actin cytoskeleton elicit changes in vacuolar structure and position, which are related to pollen hydration and germination (Iwano *et al.*, 2007). In *Brassica napus*, break-down of microtubules in stigmatic papillary cells was observed during compatible pollination (Samuel *et al.*, 2011). In addition to alterations of subcellular structures during pollination, reactive oxygen species (ROS) signalling and Ca^{2+} dynamics have also been observed in stigmatic papillar cells during the early stages of pollination (Elleman & Dickinson, 1999; Iwano *et al.*, 1999; Iwano *et al.*, 2004; McInnis *et al.*, 2006). The high level of ROS (predominately H_2O_2) in stigmas was found to be reduced after pollen adhesion, which might be caused by nitric oxide (NO) produced by the adhered pollen grain (McInnis *et al.*, 2006). As calcium is an important regulator of pollination (reviewed in Steinhorst & Kudla, 2013), it was not surprising that increased level of cytosolic free calcium was detected near the surface of both pollen and

stigma papilla cells after the application of compatible pollen and during the rejection of incompatible pollen (Iwano *et al.*, 2004; Iwano *et al*., 2015). In *Arabidopsis thaliana*, three cytosolic calcium increases were observed in the apical region of stigmatic papillar cells beneath the adhered compatible pollen grain: the first Ca^{2+} increase followed pollen hydration, the second occurred before pollen germination and the third was observed when pollen tube penetrated the stigmatic cell wall (Iwano *et al*., 2004). A recent study provided evidence of Ca^{2+} export from the papillar cell elicited only by compatible pollen coat extracted from Arabidopsis (Iwano *et al.*, 2014). Identified by transcriptomic analyses of stigma, a gene encoding an autoinhibited Ca^{2+}-ATPase (ACA13) located on the papillar cell membrane was discovered to be induced by the adhesion of both the compatible pollen grain and pollen coat. Gene expression and functional studies demonstrated that ACA13 concentrated around the pollen tube penetration site and was essential for pollen germination and tube growth, which suggested that ACA13 may function as a Ca^{2+} pump during the early stages of pollination (Iwano *et al.*, 2014).

The work on the molecular basis of SI in the Brassicaceae (see section 1.3.3.1) has shed light on the basal pollen-recognition system that operates during the early stages of pollination. The pathway of self-pollen acceptance is now thought to be overridden by the SI pathway during pollen hydration (Safavian & Goring, 2013). EXO70A1 is one of the subunits of exocyst complex that functions to tether secretory vesicles onto the plasma membrane (reviewed in He & Guo, 2009). In *Brassica*, EXO70A1 has been shown to be the substrate of ARC1, an E3 ubiquitin ligase that mediates a protein degradation pathway of SI response (Samuel *et al.*, 2009; Indriolo & Goring, 2014). The study of RFP: Exo70A1 in Arabidopsis revealed a dynamic change of RFP distributions in stigmatic papillar cell following compatible pollination. The overexpression of Exo70A1 in SI *Brassica napus* partially overcame the SI response (Samuel *et al.*, 2009). These discoveries have indicated that EXO70A1 acts at the intersection of stigmatic responses during the recognition of both self-incompatible pollen and compatible pollen, where two opposing cellular pathways control pollen hydration (Samuel *et al.*, 2009; Haasen & Goring, 2010).

1.3 Cysteine-rich proteins (CRPs) in Plants

In recent years, a broad range of small cysteine-rich proteins (CRPs) have been discovered in plants acting as important factors in pathogen defence, plant-bacterial symbiosis, development and reproductive signalling (Figure 1.6). Accumulating evidence demonstrates that they act at multiple stages of the reproductive process. CRPs are characterised by their small size (less than 160 amino acids), having a secretory signal peptide at the N-terminal and a C-terminal cysteine-rich domain (reviewed in Marshall *et al.*, 2011). Within this cysteine-rich motif, the cysteine pattern determinates the classification of this large and diverse protein superfamily (Figure 1.6). Functional studies of CRPs that have been clarified to date, show they typically act as signalling factors that regulate multiple biological activities during the life of a plant. Computational genomic research has revealed that CRPs are the most predominant secreted small peptides in *Arabidopsis thaliana* (Silverstein *et al.*, 2007). The diversity and large copy numbers of CRPs may be the result of cysteine motif rearrangement during their evolution after multiple events of WGD and small-scale gene duplications. The thermally stable structure maintained by cysteine bridges is believed to be important for the evolvability of CRPs (Sancho *et al.*, 2005; Bloom *et al.*, 2006; Silverstein *et al.*, 2007). However, intriguing and fundamental questions remain such as what the evolutionary relationships between different classes of CRPs are and how they evolved to play different roles. By reviewing the characterised CRPs involved in multiple aspects of plant life, some clues might be revealed for answering these questions.

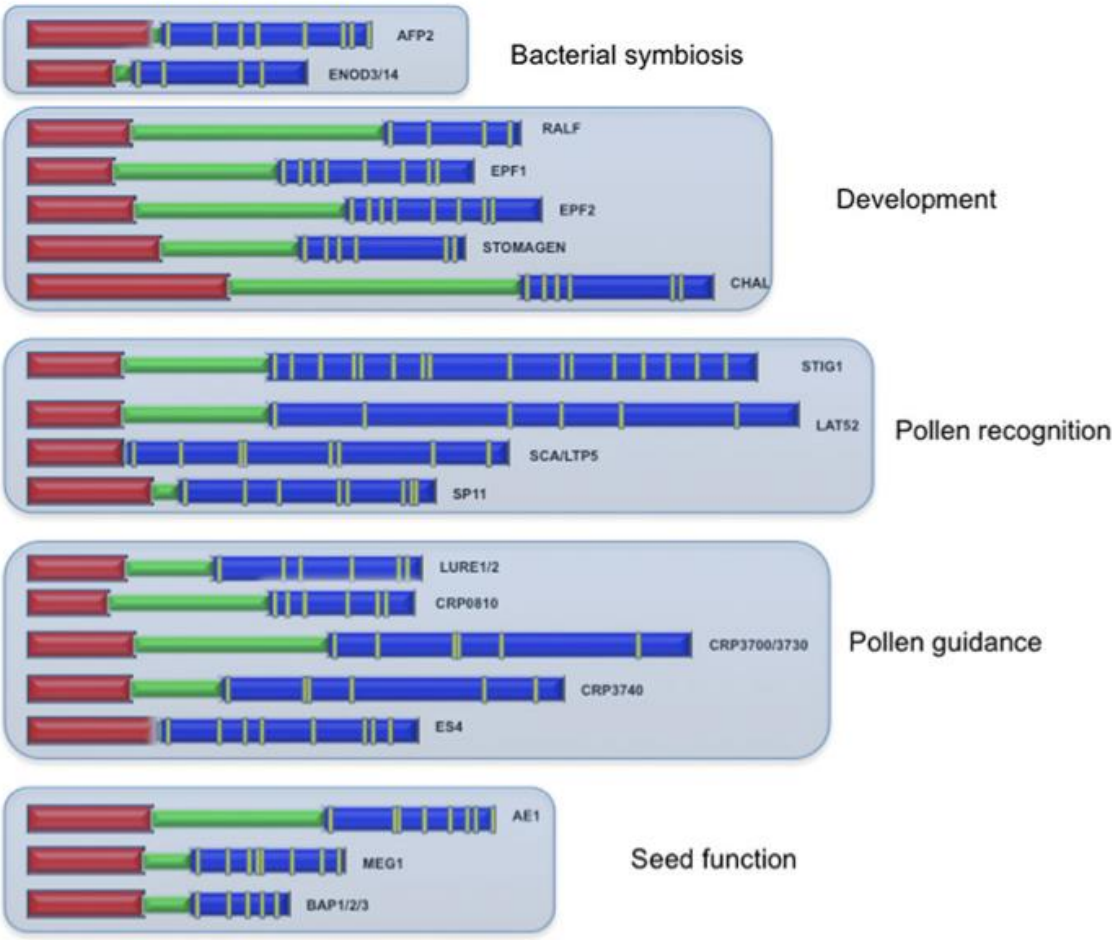


Figure 1.6 | Schematic of the conserved protein domains of plant CRPs (Marshall *et al*., 2011). Red bar, secretion signal peptide; green bar, variable region; blue bar, cysteine-rich domain with cysteine residue positions shown as yellow bars. Proteins are drawn to scale.

1.3.1 Cysteine-rich proteins in plant development

The growth and development of angiosperms, unlike animals, is largely driven by cell expansion rather than cell migration. The coordination of development from the seed to different organs through multiple pathways is crucial for survival and reproductive success. In recent years, several CRPs have been discovered to play important roles during plant development. The *Rapid Alkalinisation Factor* (*RALF*) CRP-encoding genes, have been identified in multiple species such as tobacco, Arabidopsis and sugarcane (reviewed in Murphy & De Smet, 2014). Initially identified in tobacco, RALF peptides were discovered to be involved in the regulation of root development, including root elongation and nutrient uptake (Pearce *et al.*, 2001). In Arabidopsis, one of the 34 *RALF* genes, *AtRALF23*, encodes a 138-aa precursor that is further cleaved to a 52-aa protein, which has been found to affect root and hypocotyl elongation (Srivastava *et al.*, 2009). It was recently discovered

that a receptor-like kinase FERONIA binds to RALF1, which caused the defect of RALF-induced root elongation (Haruta *et al.*, 2014). The interaction initiates a phosphorylation signalling cascade that inhibits root membrane H^+ -ATPase activity and increases apoplastic pH, which reduces cell elongation (Haruta *et al.*, 2014). Plant stomata are important for regulating air exchange, transpiration and CO_2 uptake. The development of stomata initiates from asymmetric division of a meristemoid mother cell (MMC) that produces a meristemoid and a stomatal-lineage ground cell (SLGC). The meristemoid cell then differentiates to guard cells (reviewed in Pillitteri & Torii, 2012). The stomata density and positioning are regulated by cell-cell communication, which is mediated by a signalling pathway that involves several CRPs (Shpak *et al.*, 2005). Three CRPs, Epidermal Patterning Factors 1 (EFP1), EFP2 and STOMAGEN were identified as antagonistic regulators of stomatal development and interact with a series of receptors (Shpak *et al.*, 2005; Hara *et al.*, 2009; Kondo *et al.*, 2010; Sugano *et al.*, 2010). These receptors initiate a cascade that culminates in mitogen-activated-protein kinase (MAPK) phosphorylating the transcription factor SPEECHLESS (SPCH), which results in a decrease of stomata lineage cell density (Lampard *et al.*, 2008). A recent study in *Nicotiana benthamiana* uncovered a signalling pathway where EP1 and EPF2 activate the MAPK (MPK6) and decrease SPCH level, whereas STOMAGEN increases SPCH (Jewaria *et al.*, 2013).

1.3.2 Cysteine-rich proteins in plant defence and plant-bacterial symbiosis

To defend against constant biotic attacks such as those from bacteria, fungi and insects, plants have evolved multiple defence mechanisms, including physical defence structures such as modified cell wall constitution and external cuticles, as well as, importantly, biochemical defensive responses. Among the molecules involved in plant defence, a group of highly abundant putative antimicrobial peptides (AMPs) has been found to possess the function of plant pathogen resistance. The majority of AMPs are characterised as small, secreted, cationic and cysteine-rich, including members of lipid transfer proteins, allergens, thionins, snakins, protease inhibitors and defensins (Silverstein *et al.*, 2007; Hammami *et al.*, 2009; Marshall *et al.*, 2011). Plant defensins are small peptides (45-54 amino acids) consisting of a conserved signal peptide and six or eight conserved cysteine residues that form three or four disulphide bridges. Commonly observed, these bonds maintain a cysteine-stabilised α-helix β-sheet motif (CSα/β), an α-core that containing the loop that

links the first β-strand and the α-helix, and a γ-core encompassing the hairpin loop that connect second and third β-strand (Bruix *et al.*, 1993; Fant *et al.*, 1998; Fant *et al.*, 1999; Yount & Yeaman, 2004; Yount *et al.*, 2007). Though the plant defensins show low conservation between their amino acid sequences, the cysteine residues, the CSα/β and the γ-core motif are commonly conserved. A large number of genes encoding defensin-like proteins have been identified in a series of model species: there are 317 *DEFLs* in *Arabidopsis thaliana* (Silverstein *et al*., 2005), 93 in *Oryza sativa* (Silverstein *et al*., 2007) and 778 in *Medicago truncatula* (Fedorova *et al.*, 2002; Mergaert *et al.*, 2003; Graham *et al.*, 2004; Silverstein *et al.*, 2007). Since the early 1990s, the antifungal activities of DEFLs have been revealed in a number of species (reviewed in Lacerda *et al.*, 2014). An early assay showed fungal response elicited by a defensin from radish seeds, *Raphanus sativus* antifungal protein 2 (RsAFP2), which resulted in ion changes such as Ca^{2+} influx to the fungal membrane and rapid alkalinisation that increase membrane permeability (Terras *et al.*, 1992; Thevissen *et al.*, 1996). The homologous proteins of RsAFP2 were later isolated from another four species including *Aesculus hippocastanum* (*Ah*-AMP1), *Clitoria ternatea* (*Ct*-AMP1), *Dahlia merckii* (*Dm*-AMP1) and *Heuchera sanguinea* (*Hs*-AFP1), which also exhibit antifungal activities that could be increased by inorganic ions (Osborn *et al.*, 1995). These results suggested that defensins act as ligands that interact with fungal membrane receptors, which could then modify ion channels. Recently, four AFP encoding genes, *Hc-AFP1-4*, were identified from *Heliophila coronopifolia,* a native South American Brassicaceae species (de Beer & Vivier, 2011). Despite the high similarities of their amino acid sequences, each of them displayed a different activity or divergent mechanism of action such as membrane permeabilisation and hyper-branching of the fungal cells (de Beer & Vivier, 2011). Their tissue-specific expression patterns also suggested that they play important roles natively in the protection of different structures (e.g. vegetative or reproductive tissue) or even involved in other biological processes (de Beer & Vivier, 2011). In another study, the analysis of spatial-temporal expression showed significant differences in the expression patterns of *DEFLs* in *Arabidopsis thaliana* and *Medicago truncatula*: *AtDEFLs* were dominantly expressed in inflorescences, whereas the majority of *MtDEFLs* were expressed in root nodules (Tesfaye *et al.*, 2013). These differences in expression pattern and function suggested diversified signalling pathways and roles of plant defensins and DEFLs.

The symbiosis between plants in the Fabaceae (legume family) and bacteria in the Rhizobiaceae family is a paradigm in the study of plant-bacterial symbioses. The terminal differentiation of bacteria in some symbioses is provoked and highly regulated by nodule-specific peptides from host plants by interacting with bacterial target molecules through multiple mechanisms (Farkas *et al.*, 2014). Based on the comparison of the nodule transcriptome in *Medicago truncatula* and *Lotus japonicas*, it was revealed that hundreds of genes are exclusively expressed in the terminal differentiating nodules but not in reversible nodulating structures (Mergaert *et al.*, 2003; Alunni *et al.*, 2007). Amongst these genes, there are mainly two classes of genes encoding secreted peptides that belong to either nodule-specific cysteine-rich proteins (NCRs) or glycine-rich proteins (GRPs) (Kevei *et al.*, 2002; Alunni *et al.*, 2007). Based on the updated *Medicago truncatula* genomic sequences, there are in total 600 genes encoding NCRs (Young *et al.*, 2011; Zhou *et al.*, 2013; Farkas *et al.*, 2014). They were identified as possessing 4-6 cysteine residues in a conserved pattern and to have highly divergent amino acid sequences (Mergaert *et al.*, 2003). Some of these NCRs possess the same four-cysteine residue pattern as two previously identified nodule factors (NF), early nodulin ENOD3 and ENOD14. They are exclusively expressed during the early stage of nodulation in pea roots (Scheres *et al.*, 1990). One of the cationic NCRs, NCR247, was believed to interact with multiple intracellular bacterial targets such as ribosomal proteins and chaperone GroEL (Farkas *et al.*, 2014). Interestingly, NCR247 functions to inhibit bacterial protein synthesis, which resembles the function of defensins and cationic AMPs (Farkas *et al.*, 2014). In addition, a homologous gene of *rapid alkalinisation factor (RALF*), *MtRALF1*, was identified in *Medicago truncatula*, which was upregulated during the initial stages of nodulation and may be involved in nodule development (Combier *et al.*, 2008).

1.3.3 Cysteine-rich proteins in plant reproduction

1.3.3.1 Cysteine-rich proteins in pollen-stigma recognition

Sexual reproduction in angiosperms is a strictly regulated process that involves complex communication between male and female reproductive structures and gametes. From the early stages of pollination to seed development, cysteine-rich proteins have been identified to play important roles at multiple stages of sexual reproduction (Figure 1.7). The most well studied and best-understood paradigm of cell-cell signalling during the initial stages

of pollen-stigma interaction is self-incompatibility (SI). SI has been categorised broadly into two classes based on different models of genetic behaviour: gametophytic SI (GSI) and sporophytic SI (SSI) (see section 1.2.2) (Takayama & Isogai, 2005). SI in the Brassicaceae belongs to the SSI type where the pollen incompatibility phenotype is determined by its diploid parent plant rather than its own haploid genome. The genetic basis of SSI is determined by the *S*-locus that contains three highly polymorphic genes. The *S-locus cysteine-rich* (*SCR/SP11*) gene encodes the male determinant (Schopfer *et al.*, 1999; Shiba *et al.*, 2001). SCR is a CRP and is localised to the pollen coat from where it has the opportunity to interact with the stigmatic female determinant, *S*-receptor kinase (SRK) (Takasaki *et al.*, 2000). Another *S*-locus gene, *S-locus glycoprotein* (*SLG*), is highly homologous to the SRK extracellular receptor domain but has been proved to be not essential for *S*-haplotype specificity (Takayama *et al.*, 1987; Nasrallah, 1997; Takasaki *et al.*, 2000). The interaction of SCR and SRK leads to the phosphorylation of ARC1, an E3 ubiquitin ligase that inhibits Exo70A1, which triggers pollen rejection (Stone *et al.*, 2003; Samuel *et al.*, 2009). In the Papaveraceae and Solanaceae, SI belongs to the GSI type where the pollen incompatibility phenotype is determined by its own haploid genome. In *Papaver rhoeas* the stigma *S*-determinant (PrsS) is a CRP that is secreted from stigmatic papillar cells and it contains four conserved cysteine residues (Foote *et al*., 1994). PrsS interacts with the male determinant, *Papaver rhoeas* pollen *S*-determinant (PrpS), a transmembrane protein expressed in pollen (Wheeler *et al.*, 2009). The interaction of *S* proteins from the stigma and the incompatible pollen grain triggers calcium influx at the pollen tube tip and then programmed-cell death (PCD) that inhibits pollen tube growth (Franklin-Tong *et al.*, 2002; Thomas & Franklin-Tong, 2004).

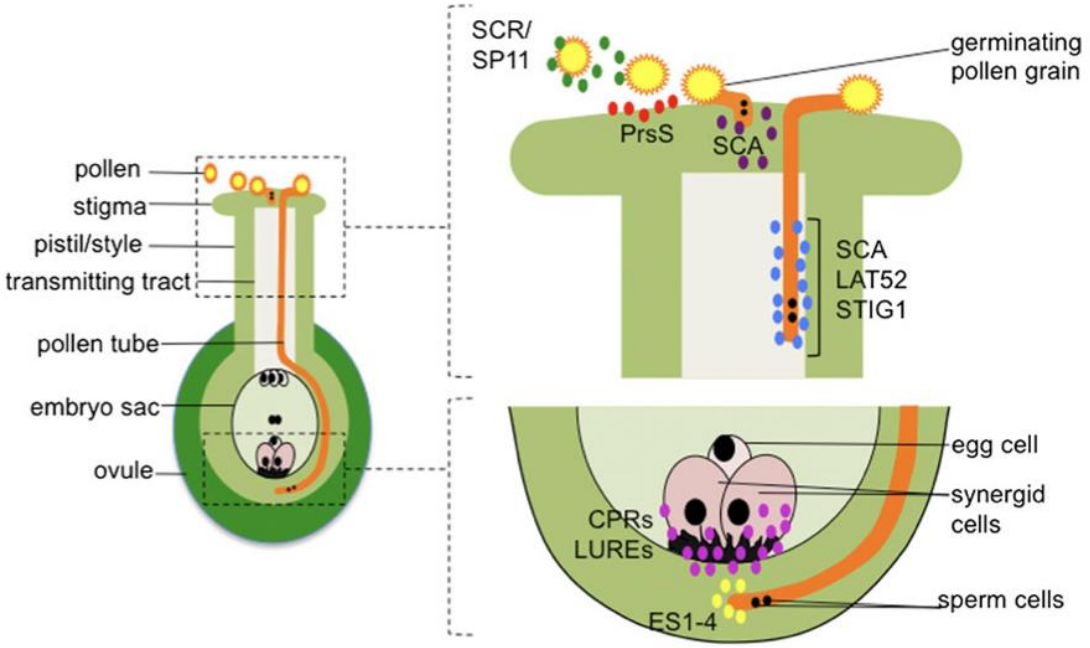


Figure 1.7 | Schematic of double fertilisation in angiosperms and cysteine-rich proteins (CRPs) involved in regulations of plant reproduction (Marshall *et al*., 2011). Left diagram represents a pollinated pistil and ovary. Top right diagram represents CRPs acting during early stages of pollen-stigma interactions including pollen hydration, germination, tube penetration and elongation. Bottom right diagram shows CRPs involved in the gametophytic interactions prior to fertilisation. The different coloured ellipses represent different CRPs.

1.3.3.2 Cysteine-rich proteins in pollen tube growth and guidance

After being recognised and accepted by the stigma, the compatible pollen grain hydrates and produces a pollen tube that grows and penetrates the stigma (as described in 1.2.1). To enter the ovary and deliver the sperm cells, the pollen tube needs to be guided towards the ovule during its growth. This directional growth requires a complex series of interactions between the pollen tube and female reproductive tissues (reviewed in Higashiyama & Takeuchi, 2015). The regulation of this process has been discovered to involve CRPs at multiple steps (Figure 1.7). In *Lycopersicon esculentum* (tomato), a small, pollen-specific CRP, Lat52, was found essential for pollen germination *in vitro* and fertilisation *in vivo* (Muschietti *et al.*, 1994). Two leucine-rich repeat (LRR) receptor kinases were characterised as *Lycopersicon esculentum* pollen-specific receptor-like kinase 1 and 2 (*Le*PRK1 and *Le*PRK2), and were identified as the receptors for Lat52. *Le*PRK1 interacts with Lat52 before germination, whereas *Le*PRK2 extracellular domain binds to Lat52 during pollen tube growth (Tang *et al.*, 2002). A pistil-specific CRP, *Le*STIG1, has also been found to interact with *Le*PRK2 (Tang *et al.*, 2004). *In vitro* pollen tube growth assays indicated that a low concentration of *Le*STIG1 was sufficient to promote pollen tube

growth (Tang *et al.*, 2004). The treatment of mature pollen extracts with exogenous *Le*STIG1 abolished the interaction of Lat52 and *Le*PRK2. These results implied that *Le*STIG1 act as a positive regulator by replacing Lat52 in interacting with LePRK2 during pollen tube growth after pollen germination on the stigma (Tang *et al.*, 2004). However, there is no evidence for supporting the role of *Le*STIG1 as a pollen tube growth regulator *in vivo*.

To reach the ovule, the pollen tube has to grow in the extracellular matrix (ECM) and to be guided alongside the transmitting tract (TT) by adhering to the TT epidermis (TTE) (Jauh *et al.*, 1997). In *Lilium longiflorum*, the *in vitro* pollen tube adhesion assays reconstructing pollen tube guidance illustrated that a small, secreted CRP facilitates pollen adhesion to the TTE (Figure 1.7) (Park *et al.*, 2000). This CRP, termed as Small Cysteine Adhesion (SCA), belongs to the lipid transfer protein (LTP) and consists of eight conserved cysteine residues and is specifically expressed in the stigma and style of lilies (Mollet *et al.*, 2000). In addition to having to growing in the right direction, a pollen tube has to be attracted to the entrance of the embryo sac and burst to release the sperm cells. In carpels having more than one embryo sac, the guidance mechanism ensures that only one pollen tube can enter each embryo sac. In *Arabidopsis thaliana*, the gamete fusion event was found to be essential for preventing polytubey (Beale *et al.*, 2012). Down-regulated expression of an egg cell-secreted CRP, EC1 protein, was found to result in delayed gamete fusion and polytubey (Figure 1.7) (Rademacher & Sprunck, 2013). In *Torenia fournieri*, two six-cysteine CRPs termed LURES have been recently discovered to be expressed in synergid cells and identified as signalling molecules that are essential for pollen tube attraction (Figure 1.7) (Okuda *et al.*, 2009). Later studies of a cluster of *DEFL* genes in *A. thaliana* identified two functional AtLURE proteins that also function to attract pollen tubes to the micropyle by interacting with the heterodimer male receptor MDIS1-MIK (Takeuchi & Higashiyama, 2012; Wang T *et al.*, 2016). Genome-wide expression analysis of Arabidopsis revealed over 50 CRPs that are exclusively expressed in synergid cells (Jones-Rhoades *et al.*, 2007). A synergid-specific transcription factor, MYB98, was found to regulate these CRPs in Arabidopsis. The fertilisation defect in *myb98* mutants demonstrated the potential function of these synergid-specific CRPs as signalling molecules that mediate cell-cell communication between the pollen tube and target female gametophyte (Punwani *et al.*, 2007; Punwani *et al.*, 2008). Moreover, there is evidence showing that pollen tube rupture also involves CRPs. In *Zea mays*, the treatment of pollen

tubes with a low concentration of the maize CRP, *Zm*ES4 (*Zea maize* embryo sac), rapidly led to the rupture of the pollen tube tip, which is likely caused by activation of the potassium import channel KZM1 (Amien *et al.*, 2010).

1.3.3.3 Cysteine-rich proteins in seed development

Double fertilisation marks the initiation of the next sporophytic generation in angiosperms and results in the development of the zygote and endosperm. From embryogenesis to the development of embryo and endosperm, accumulating evidence showed the importance of CRPs in cell-cell communication and signalling. In Maize, wheat and Arabidopsis, transcriptomic analyses revealed CRP expression in female gametes (Dresselhaus *et al.*, 1994; Le *et al.*, 2005; Sprunck *et al.*, 2005; Yang *et al.*, 2006). Utilisation of a laser-assisted microdissection (LAM) enabled cell type-specific transcriptomes to be obtained for the egg, the central cell and the synergid cells in Arabidopsis. There were 33 DEFLs presented on the microarray and seven of them were highly enriched in the female gametophyte (Wuest *et al.*, 2010). The development of embryo and endosperm requires the coordination of their growth with the surrounding tissue, which may involve a series of CRPs in Arabidopsis. These CRPs are expressed in the female gametes as well as in the developing zygote and endosperm (Steffen *et al.*, 2007). Exclusive expression of CRPs in different domains of endosperm has been observed. In maize, some CRPs are expressed in the embryo surrounding region (ESR) (Magnard *et al.*, 2000; Balandin *et al.*, 2005), others are expressed in the basal endosperm transfer layer (BETL) cells, such as Maternally Expressed Gene 1 (MEG1) (Gutierrez-Marcos *et al.*, 2004) and Basal layer Antifungal Proteins (BAPs) that interestingly exhibit antifungal activities (Serna *et al.*, 2001). The homologues of *MEG1* in Arabidopsis encode a family of small, cysteine-rich proteins termed as EMBRYO SURROUNDING FACTOR 1 (ESF1). ESF1 proteins are exclusively expressed in the central cell and ESR and act as positive regulators of early development of suspensor and zygote elongation (Costa *et al.*, 2014).

Despite the diverse of CRP proteins in plants, it is not difficult to notice some striking structural and functional similarities amongst classes of this protein superfamily. CRPs play significant roles particularly in recognition signalling, such as plant-pathogen interaction, plant-bacteria symbiosis and reproductive cellular communication. The similarity of cysteine patterns between some CRP classes, such as DEFLs and PCP-As,

implicated a possible common evolutionary origin of the genes encoding these signalling proteins. Although at the moment it is not clear that how these CRPs evolved their diverse functions, the ongoing structural and evolutionary studies may provide some further clues. In particular, large amount of studies have been focusing on the identifying of ligand-receptor pairs in plants (see Chapter 4, 4.1, reviewed in Endo *et al*., 2014), which has been providing evidence showing that CRPs are important ligand molecules in receptor-mediated signalling.

1.4 Pollen coat and cysteine-rich Pollen Coat Proteins (PCPs)

In the Brassicaceae, including *Brassica oleracea* and *Arabidopsis thaliana*, the pollen cell surface contains three strata: the outermost layer is the pollen coat (tryphine), then a multi-layered outer exine wall that consists of sporopollenin and incorporates apertures, and finally underlying this is the intine which is made of cellulose. The pollen coat is a complex mixture that fills in the pollen exine cavities and it contains lipids, proteins, glycoconjugates and pigments (Figure 1.8) (Heslop-Harrison, 1968; Piffanelli *et al.*, 1998; Hernandez-Pinzon *et al.*, 1999). During pollen development, the tapetal cells which line the anther locule lyse and the contents are transferred to the pollen grain surface to form the pollen coat (Dickinson & Lewis, 1973). Although several molecular regulators of compatibility have been identified in stigmas, less is known about the pollen-borne factors. Analyses of pollen coat components and mutational studies in the Brassicaceae have revealed factors influencing the pollen-stigma interaction. However, most of these appear to influence the biophysical properties of the pollen coat. For example, as mentioned previously (1.2.2), the elimination of very long chain lipids from the pollen coat in *eceriferum* (*cer*) mutants leads to the failure of pollen hydration (Preuss *et al.*, 1993; Hulskamp *et al.*, 1995; Fiebig *et al.*, 2000). Rather than acting as recognition molecules, these lipids are likely playing a role in solubilising and mobilising other factors by establishing hydration conduits (Elleman *et al.*, 1992; Dickinson, 1995; Hulskamp *et al.*, 1995; Wolters-Arts *et al.*, 1998; Edlund *et al.*, 2004). Impaired pollen hydration has also been observed in mutants lacking extracellular lipase 4 (EXL4) and an oleosin domain-containing glycine-rich protein (GRP17) (Mayfield & Preuss, 2000b; Updegraff *et al.*, 2009). Similarly, it is likely that EXL4 and GRP17 act as modifiers of the lipid

composition at the pollen-stigma interface where water passage can be established for pollen hydration.

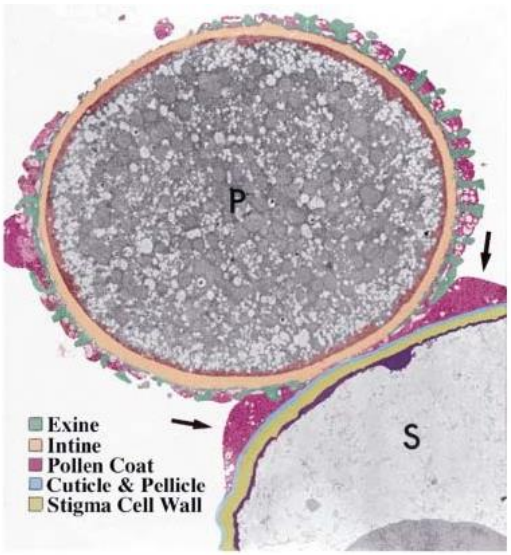


Figure 1.8 | Transmission electron microscopy (TEM) showing the contact point of a pollen grain (P) interacting with a stigma papillar cell (S) in Arabidopsis. The TEM image is artificially coloured for highlighting the pollen coat (pink), intine (peach), exine (green), stigma cell wall (yellow) and stigma cuticle (blue). The arrows are indicating the lipid-rich 'foot' between the two surfaces (Edlund *et al.*, 2004).

As discussed previously, CRPs have been revealed to have important functions in plant reproduction. Several families of CRPs have been identified in the pollen coat of *Brassica* species and Arabidopsis (section 1.3.3.1). Based on the numbers of cysteine residues and their arrangement (cysteine pattern), these pollen coat CRPs can be divided into three classes, the SCRs, PCP-As and PCP-Bs (Doughty *et al.*, 2000). The SCRs, (see section 1.3.3.1), include the male determinant of SI in *Brassica* (Schopfer *et al.*, 1999; Takasaki *et al.*, 2000; Shiba *et al.*, 2001). Two PCP-A class pollen coat CRPs have been characterised to date. SLR-binding protein (SLR-BP1) identified in *Brassica campestris*, interacts with the *S*-locus related (SLR1) stigmatic protein and may play some role in pollen adhesion (Takayama *et al.*, 2000). Another member of the PCP-A class, PCP-A1, was identified in *Brassica oleracea* pollen coat, and binds the stigmatic *S*-locus glycoprotein (SLG) (Doughty *et al.*, 1998), however its function is unknown. Similar with the PCP-As, the members of PCP-B class also possess a secretory signal peptide and a cysteine-rich region containing eight cysteine residues, which however are arranged differently (Figure 1.9). Protein sequencing of *Brassica oleracea* pollen coat proteins revealed two PCP-Bs (*Bo*PCP-B1 and 2) but again their function was not determined (Doughty *et al.*, 2000). Accumulating evidence indicates that the pollen coat CRPs are important signalling

molecules in the pollen-stigma interaction. Previous studies were focused on the SI system in a range of families, but the very fundamental regulation of basal compatibility in angiosperms is still poorly understood.

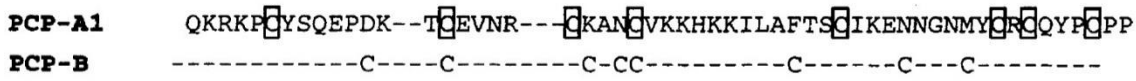


Figure 1.9 | Cysteine-rich region alignment of PCP-A1 and PCP-B class proteins (Doughty *et al.*, 2000).

1.5 Aims and objectives

The initial stages of the pollen-stigma interaction are tightly regulated by a basal compatibility system established by a molecular dialogue between pollen and stigma. Despite its fundamentally important role in intraspecific pollination and reproductive isolation, very little is known about the molecular basis of this recognition system. This project aimed to identify and characterise potential pollen ligands and stigmatic receptors involved in the early stages of pollen-stigma compatibility. The results provide new insights into the molecular regulation in plant reproductive signalling and evolutionary significance of these molecules.

The project aims detailed above were accomplished via the following research objectives:

- Accumulating evidence has demonstrated that the pollen coat carries important regulatory factors that mediate pollen-stigma interactions. Cysteine-rich proteins (CRPs) play important roles in regulating successful fertilisation in plants. A novel class of cysteine-rich pollen coat proteins (PCPs), the PCP-Bs, was previously identified in *Brassica oleracea.* Due to the availability of T-DNA insertion gene knockout lines in *Arabidopsis thaliana*, functional analyses of the PCP-Bs were carried out as part of this project. Specifically four Arabidopsis PCP-Bs were identified to be expressed gametophytically late in pollen development. Bioassays of early-stage pollen-stigma interactions were performed by utilising T-DNA gene knockout lines to address the functions of each of these PCP-Bs.

- Previous studies on pollen coat proteins involved in self-incompatibility system in Brassica suggested that these pollen coat CRPs are acting as signalling molecules and interacting with stigmatic targets to mediate both compatible and incompatible pollen-pistil interactions. The search for stigmatic receptor(s) of PCP-Bs is intriguing and essential for furthering our understanding of the molecular basis of pollen-stigma recognition system in angiosperm sexual reproduction. In this study, four AtPCP-Bs were heterologously expressed and utilised for identifying potential stigmatic targets. Various protein-protein interaction assays were carried out with these AtPCP-B ligands acting as 'bait'. A number of stigmatic proteins were identified as putative AtPCP-B receptors and future work is necessary to validate these preliminary candidates.

- Reproductive isolation mediated by sexual reproductive barriers plays important roles in the rapid diversification of angiosperms. Previous evidence suggested gamete recognition proteins are more likely to undergo adaptive evolution since the coevolution between the male and female recognition proteins could perform positive selective force to each other. The BLAST sequences searches through currently available plant genomes revealed large number of highly diverged sequences of PCP-Bs and putative homologues. This sequence resource was utilised to analyse the evolutionary history and molecular evolution of this novel pollen coat CRP family. In this project, phylogeny analysis of gene sequences encoding PCP-Bs and their homologous proteins provided preliminary evidence of rapid diversification and multiple gene duplication events. Molecular evolutionary analysis showed evidence of positive selection on the sites of PCP-Bγ and its orthologues, which provided further evidence to support the function of PCP-Bs in pollen-stigma recognition.

- The pollen coat has been found to carry important factors that mediate the early stages of pollen-stigma interaction. By being localised at the interface of pollen and stigma contact, pollen coat components are rapidly released to engage in intercellular molecular communication. Although previous profiling of the pollen coat in several species has uncovered a range of pollen coat protein components, the rapidly developing sensitive and cost-effective detection techniques can now be utilised to the

characterisation of scarce samples such as Arabidopsis pollen coat preparations. In this study, liquid chromatography/mass spectrometry (LC-MS/MS) was utilised to characterise the full protein complement of pollen coat from Arabidopsis and *Brassica oleracea*. The proteomic profiling of pollen coat from both species revealed an astonishingly large number of proteins that have not been previously reported. The strikingly number of CRPs in the data also provides a precious source of molecules that can be studied for their potential roles in the basal compatibility system in the pollen-stigma interactions.

Chapter 2 Materials and methods

2.1 Plant materials

2.1.1 Plant growth conditions

Arabidopsis thaliana plants were propagated in Levington F2+S compost (Soils HS Limited, Wotton-Under-Edge, UK) in a controlled environment room with a 16h: 8h light: dark photoperiod provided by fluorescent lighting (130 μmol $m^{-2}sec^{-1}$). The temperature was maintained at 21±1°C with 60% relative humidity. Brassicas were grown in a glasshouse at 21°C with a 16h: 8h light: dark photoperiod.

2.1.2 Arabidopsis thaliana lines

Arabidopsis thaliana (L.) Heynh. T-DNA insertion lines SALK_207087, SALK_062825, SALK_072366 (Alonso *et al.*, 2003) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). GABI_718B04 was purchased from GABI-KAT (Kleinboelting *et al.*, 2012). T-DNA insertion sites and their respective mutant alleles are detailed in Fig. S1.1. Single gene T-DNA insertion lines were backcrossed to wild-type (Col-0) at least three times before phenotyping. *pcp-bβ/γ* and *pcp-bα/β/γ* mutant lines were created using standard crossing procedures. *PCP-B* transcript status for each mutant was confirmed by reverse transcription polymerase chain reaction (RT-PCR) (Appendix 1, Fig. S1.2). Primers for RT-PCR are detailed in Table S1.1. The A9-barnase male sterile line was provided by Rod Scott, University of Bath, UK (Paul *et al.*, 1992). GUS (β-glucuronidase) reporter lines *pAt5g61605:GUS* and *pAt2g16505:GUS* were provided by José F. Gutierrez-Marcos (University of Warwick, UK).

2.2 Genotypic analyses of T-DNA insertion lines

2.2.1 DNA extraction

The genomic DNA used in this project was extracted from fresh leaves of *Arabidopsis thaliana*. Approximately 0.5cm^2 of Arabidopsis leaf was cut and ground in 100µl ice-cold extraction buffer (0.2 M Tris pH 8.0, 0.1 M EDTA, 1% Sarkosyl, 100 g ml^{-1} proteinase K). Incubate the mixture at 65°C for five minutes. 100µl of chloroform was added and vortexed with the mixture for 30 seconds. The mixture was centrifuged (5415D, Eppendorf) at 13,000***g*** for 10 minutes. The upper phase was transferred to a clean 1.5 ml microcentrifuge tube. DNA was precipitated by adding 100µl of isopropanol, mixing gently and incubating at room temperature for 15 minutes. The sample was centrifuged at 13,000***g*** for 20 minutes to pellet DNA. The supernatant was removed and the pellet was incubated in a 37°C heat block until no liquid could be observed. The DNA pellet was resuspended in 50µl of milliQ water.

2.2.2 Genotyping and phenotyping of T-DNA gene mutants

The genotypes of gene knockout lines were confirmed by the PCR of genomic DNA (leaf) with primers recommended by T-DNA Express (http://signal.salk.edu/cgi-bin/tdnaexpress) primer design tool (iSect Primers). PCR was carried out with DreamTaq Green PCR Master Mix (2X) in the MJ Research PTC-200 Thermal Cycler (GSI Inc, Minnesota, USA). PCR primers used for T-DNA line confirmation can be found in Appendix 1 Table S1.1-C**.**

2.2.3 The T-DNA insertion location confirmation by DNA sequencing

The PCP-B gene transcript status of each mutant was confirmed by RT-PCR. PCR primers used for RT-PCR can be found in Appendix 1 Table S1.1-B. The T-DNA insertion location was confirmed by DNA sequencing.

2.3 Transcriptional analysis by reverse transcription polymerase chain reaction (RT-PCR)

2.3.1 RNA extraction

Anthers were collected from *A. thaliana* flower buds at stage 10-12 (Smyth *et al.*, 1990), stigmas from open flowers of the *A. thaliana* A9-barnase line, roots from two-week-old seedlings grown on ½ MS plates and leaves from fully-grown rosettes. All the plant tissue samples were flash frozen in liquid nitrogen in 1.5 ml microfuge tubes and were stored at -80°C as soon as they were dissected from the living plant. RNA was extracted using a PureLink® RNA Mini Kit (Thermo Fisher Scientific).

2.3.2 cDNA synthesis

The synthesis of cDNA was carried out by using ProtoScript® II First Strand cDNA Synthesis Kit (New England Biolabs) in the MJ Research PTC-200 Thermal Cycler. The reactions were stored at -20°C or used for PCR immediately.

2.3.3 Reverse transcription polymerase chain reaction (RT-PCR)

Phusion® High-Fidelity DNA Polymerase (New England Biolabs) was used for RT-PCR followed by cloning reactions for guaranteeing high fidelity of base pairs. DreamTaq® Green PCR Master Mix (2X) (Thermo Fisher Scientific) was used for routine diagnostic PCR Primers used for DNA amplification can be found in Appendix 1 Table S1.1. The positive control for RT-PCR reactions was carried out with *GapC* gene (*At3g04120*) specific primers to prove the cDNA synthesis was successful. Genomic DNA used as a positive control was extracted from leaves of the *Arabidopsis thaliana* (Col-0) (See DNA extraction in 2.2.1). 1µl of cDNA or genomic DNA was used in each reaction. The same amount of DEPC-treated water was used as a negative control to exclude the possibility of contamination.

2.3.4 PCR product verification with electrophoresis.

The PCR products were run on 1% agarose gel electrophoresis at 120V for 30 minutes.

2.4 Functional study of AtPCP-Bs

2.4.1 Pollen hydration assay

For *in vivo* hydration assays, pollen grains derived from Col-0, *pcp-bα* (SALK_207087), *pcp-bβ* (SALK_062825), *pcp-bγ* (SALK_072366), *pcp-bδ* (GABI_718B04), *pcp-bβ/γ* and *pcp-bα/β/γ* lines were applied to stigmas of the A9-barnase male sterile line. Freshly opened mature flowers were used (retained on the plant) and pollen was applied in a monolayer with an eyelash. At least eight independent stigmas were used in assays from each line. For *in vitro* hydration assays pollen from Col-0 and *pcp-bα/β/γ* triple knockout lines were placed on a slide in a humid chamber (100% relative humidity). Pollen grains on stigmas were photographed under a dissecting microscope immediately after pollinations were initiated (designated as time point zero). Subsequent images were captured every minute for 30 minutes. For humid chamber assays, pollen was photographed one minute after pollen grains were placed, then the chamber was sealed, after which images were taken every minute for 30 min. Equatorial diameter of pollen was measured in pixels using ImageJ Software (Schneider *et al.*, 2012). Pollen hydration (%) was calculated using the equation: Pollen hydration (%) = (Pollen diameter – initial pollen diameter)/ initial pollen diameter. Slopes were determined using 11 data points during the 0-10 min, 10-20 min, or 20-30 min time periods using the linear regression curve $f = a_0 x + b$. All statistical analyses were carried out using Microsoft Excel 2013.

2.4.2 Pollen adhesion assay

Stigmas of *A. thaliana* A9-barnase plants were hand-pollinated using freshly dehiscent anthers from Col-0 and *pcp-bα/β/γ* lines. Pollen grains were applied as a monolayer. After 30 minutes the flower was excised from the plant and placed into 0.5ml of fixative (60% v/v ethanol, 30% v/v chloroform, 10% v/v acetic acid) in a 1.5ml microfuge tube. The sample was immediately shaken 10 times using short sharp strokes to dislodge pollen that was not strongly adherent to the stigma. The flower was then removed and placed into a separate microfuge tube. Both samples were retained for pollen counting. 50 µl of aceto-orcein stain (1%) was added to the tubes and incubated overnight at room temperature (RT). 'Washed-off' pollen samples were centrifuged at 13,000$\boldsymbol{g}$ (10 minutes), excess

fixative was removed and the pellet was resuspended in 10 µl of 50% glycerol before counting. Stigmas were excised from stained flowers before mounting in 50% glycerol and were squashed on a slide to ensure all pollen was visible for counting.

2.4.3 Pollen tube growth assay

Pollinations were initiated on *A. thaliana* male sterile A9-barnase stigmas and allowed to proceed for two or four hours before stigmas were excised and incubated in fixative (60% v/v ethanol, 30% v/v chloroform, 10% v/v acetic acid) overnight. After removal of fixative, stigmas were incubated in 8M NaOH for 20 minutes then washed in dH_2O 3 times, each for 5 minutes. Samples were transferred to 0.1% decolourised aniline blue (0.1% w/v aniline blue in 0.1M K_3PO_4, pH 11) for 1 hour before imaging (Kho & Baer, 1968).

2.5 Microscopy

2.5.1 Optical microscopy

Imaging of pollen hydration on stigmas, hydration in a humid chamber and GUS histochemical staining (Appendix 1 Method S1.1) of flowers and leaves was carried out using a Nikon SMZ1500 dissection microscope coupled to Nikon Digital Sight DS-U1 camera. Anther and stigma collecting was carried out under a dissection microscope (Leica MZ6). The pollen counting of the pollen adhesion assay was carried out using a compound microscope (Olympus BH-2). A Nikon Eclipse 90i epifluorescence microscope (10 X objective) with Nikon Digital Sight DS-U1 camera was used for imaging pollen tubes stained with aniline blue and for anthers stained for GUS activity.

2.5.2 Scanning and Transmission Electron Microscopy

Flowers of the male sterile Arabidopsis Col-0 A9-barnase line were pollinated with pollen grains derived from the mutant *PCP-B* lines *pcp-bα* (SALK_207087), *pcp-bβ* (SALK_062825), *pcp-bγ* (SALK_072366), *pcp-bδ* (GABI_718B04) knockout lines, *pcp-bβ/γ* double knockout line and *pcp-bα/β/γ* triple knockout line. Pollinated stigmas were dry-fixed using the method of Elleman and Dickinson (1986). Samples were washed with 0.1M Sodium Cacodylate Buffer pH 7.4 and prepared for SEM and TEM by a modified method

of Villar *et al*. (1987). Pre-fixation was completed by an overnight treatment of 2.5% glutaraldehyde in 0.1M SCB pH 7.4. After washing with 0.1M SCB, the samples were post-fixed using 1% osmium tetroxide in 0.1M SCB pH 7.4 for two hours at room temperature. The samples were then washed in 0.1M SCB pH 7.4 followed by 0.05M Maleate buffer pH 6.0. Samples were stained by 1% Uranyl Acetate in Maleate buffer pH 6.0 for 1 hour in the dark and washed by Maleate buffer pH 6.0. The samples were dehydrated in an acetone series. Samples for SEM were dried using a Polaron E3000 Critical Point Dryer (Agars Scientific UK) in CO_2, coated with gold and imaged using a JEOL JSM640LV Scanning Electron Microscope (JEOL Tokyo Japan). Samples for TEM were embedded with Spurr's Epoxy Resin Kit. Ultra-thin sections were produced as 100 nm on Ultracut-E ultramicrotome (Leica UK). Samples were imaged by JEOL JEM1200EXII Transmission Electron Microscope.

2.5.3 Histochemical staining for β-glucuronidase (GUS) activity

Leaves, open flowers, stage 12 buds of *pAt5g61605: GUS* and *pAt2g16505: GUS* were transferred into GUS substrate solution (100 mM sodium phosphate buffer pH7.0, 10mM EDTA, 0.1%(v/v) Triton X-100, 1mM potassium ferricyanide and 2mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronide) and vacuum-infiltrated for 5 minutes before incubation overnight at 37°C. Samples were then transferred into 50% (v/v) ethanol to remove chlorophyll pigmentation after GUS staining. The samples were mounted with 50% (v/v) glycerol for imaging.

2.6 Bioinformatics

2.6.1 Multiple sequence alignment and phylogenetic analysis

PCP-B-like (*PCPBL*) sequences were retrieved from available complete plant genomes completed to at least scaffold level using TBLASTN database searches (Phytozome, https://phytozome.jgi.doe.gov; Comparative Genomics, CoGe, https://genomevolution.org). Nucleotide sequence alignments of *PCP-B* homologous genes were generated using MUSCLE (codon) (Edgar, 2004). Codons of protein coding sequences were translated into amino acid sequences before the alignment was performed. Aligned amino acid sequences

were then replaced by the original codons. Graphical output of protein sequence alignment was generated by Jalview using 'ClustalX' colour coding. Phylogenetic trees were built using the Maximum Likelihood statistical method in MEGA (version 6) (Tamura *et al.*, 2013). The initial tree was determined by neighbour-joining method (NJ). The phylogeny test was carried out using the bootstrap method (1000 replications). Phylogenetic trees were displayed using iTOL (Letunic & Bork, 2007).

2.6.2 Estimation of the gene gain and loss events

To infer the gene duplication and loss events of PCP-B-like (PCPBL) proteins and Ubiquitin-like 5 proteins in 15 plant species, the gene trees and species tree were reconciled by using Notung (version 2.8) (Chen *et al.*, 2000).

2.6.3 Conserved motif analysis and gene syntenic analysis

The conserved motifs in PCP-B-like (PCPBL) protein sequences (signal peptides were excluded) in 15 plant species were identified by using MEME programme (version 4.11.1) (Bailey *et al.*, 2009). The local run was carried out on web server (http://meme-suite.org) with following parameters: any number of repetitions in each sequence, maximum number of motifs = 10, minimum optimum motif widths: 6 residues, minimum optimum motif widths: 50 residues.

2.6.4 Maximum Likelihood tests of positive selection

The ratio (ω) of nonsynonymous to sysnonymous substitution rates (d_N/d_S) was used as an indicator of selective pressures acting on the CRP coding region (the signal peptides were ignored). To analyse the d_N/d_S value of a pair of genes, the YN00 tool in the PAML package (v4.8) was used (Yang & Nielsen, 2000). To analyse the branch specific d_N/d_S value on a phylogeny, the branch model (free-ratio) in the codeml tool in the PAML package was carried out. The site-specific selection analysis was processed using the codeml tool in the PAML package (v4.8). The heterogeneity of ω on branches was evaluated by comparing free-ratio model and one-ratio model (M0). Another three pairs of models were used to test whether some sites are under positive selection (Table 2.1):

(1) The M1a (Nearly neutral) and M2a (Positive selection);
(2) The M7 (beta) and M8 (beta&ω);
(3) The M0 and M3 (discrete).

The model comparisons were performed using likelihood ratio test (LRT). The LTR statistic follows Chi-square distribution. The number of additional parameters in the more complex model determines the degree of freedom (*df*). The critical value based on a certain *df* can be obtained from standard statistic tables. Twice the log-likelihood ($2\Delta \mathrm{In}L$) difference between two models was compared to the critical value and the ***p***-value (level of significance) was calculated based on *df* and $2\Delta \mathrm{In}L$.

Tabel 2.1| Parameters in the site models

Model	NSsites	#p[f]	Parameters	References
M0 (one ratio)[a]	0	1	ω	(Goldman & Yang, 1994; Yang & Nielsen, 1998)
M1a (neutral)[b]	1	2	$p_0(p_1=1-p_0)$, $\omega_0<1$, $\omega_1=1$	(Nielsen & Yang, 1998; Yang *et al.*, 2005)
M2a (selection)[c,g]	2	4	$p_0, p_1(p_2=1-p_0-p_1)$, $\omega_0<1$, $\omega_1=1$, $\omega_2>1$	(Nielsen & Yang, 1998; Yang *et al.*, 2005)
M3 (discrete)	3	5	$p_0, p_1(p_2=1-p_0-p_1)$, ω_0, ω_1, ω_2	(Yang, 2000)
M7 (beta)[d]	7	2	p, q	(Yang, 2000)
M8 (beta&ω)[e,g]	8	4	$p_0(p_1=1-p_0)$, p, q, $\omega_2>1$	(Yang, 2000)

a. There is only one ω value for all branches.
b. It allows sites to fall into two catergories, ω<1 and ω=1.
c. It allows sites to fall into three catergories, ω<1, ω=1 and ω>1.
d. It allows ten ten grades of ω sites between 0 and 1 based on a beta distribution with parameters p and q.
e. It adds an additional grade of ω>1 as M2a.
f. The number of free parameters in the ω distribution.
g. The candidate sites for positive selection were indicated by using Bayes Empirical Bayes (BEB) analysis.

2.6.5 Protein structural prediction and modelling of AtPCP-Bs

The protein sequence alignment of PCP-Bs and 2RU1 was carried out with T-COFFEE (Notredame *et al.*, 2000). The PCP-B protein models were built by SWISS-MODEL (Arnold *et al.*, 2006; Guex *et al.*, 2009; Kiefer *et al.*, 2009; Biasini *et al.*, 2014) based on the modified T-COFFEE alignment result. The three-dimensional cartoon models and electrostatic potential surface models were produced by PyMOL (version 1.7.4).

2.6.6 Gene ontology

The mass spectrometry data set of Arabidopsis pollen coat profiling was searched against the PANTHER (**P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships) Classification System GO-slim (Version10.0), which describes the functions of gene products.

2.6.7 Gene syntenic analysis

Gene syntenic analysis was carried out by using GEvo

(https://genomevolution.org/coge/GEvo.pl).

2.7 Gene cloning

2.7.1 Restriction enzyme digestion and DNA ligation

PCR and digested products were recovered from electrophoresis agarose gels by using PCR purification system Wizard® SV Gel and PCR Clean-Up System (Promega). The target genes were cloned into cloning vector pGEM®-T Easy Vector (Promega) or pJET1.2 cloning vector from CloneJET PCR Cloning Kit (Thermo Fisher Scientific). To express the recombinant *AtPCP-Bs* in *E.coli*, the cloned genes were cloned into expression vector pET-32a (Novagen, Appendix 2.3). The digested products were analysed and separated on 1% electrophoresis agarose gel. Purified digested target DNA and vector were ligated by T4 DNA ligase (Promega).

2.7.2 Transformation

The recombinant cloning vectors were transferred into JM109 High-Efficiency Competent Cells (Promega) and the recombinant expression vectors were transferred into Origami 2 (DE3) competent cells (Novagen). The transformation was carried out by following the protocol of pGEM®-T Easy Vector System I (Promega).

2.7.3 Colony PCR and DNA sequencing

The transferred colonies were analysed by colony PCR to detect the presence of inserted target genes. The colony PCR was carried out with T7 promoter forward primer and SP6 reverse primer or with gene specific primers. The recombinant vectors were extracted from harvested cell culture by GeneJET™ Plasmid Miniprep Kit (Thermo Fisher Scientific). The extracted recombinant vectors were verified by DNA electrophoresis and sequencing tests (Sanger Sequencing Service, Source BioScience) were carried out for confirming gene sequences and open reading frames (ORF).

2.7.4 Bacteria culture growth and storage

Liquid LB Media (Appendix 2.2.1) was sterilized by autoclaving at 121°C for 20min before use. Solid LB media with 2% (w/v) agar (Bacto-agar) was autoclaved and kept warm before making agar plates. Antibiotics stocks were filtered through a 0.22μm filter

(Millipore) and were added into the media when the media temperature was below 55°C. *E.coli* was cultured at 37°C in LB medium routinely.

For short-term storage, plates with cultures were sealed and stored at 4°C. For long-term storage, the cultures were prepared as glycerol stocks: the fresh overnight-incubated 5ml culture was centrifuged and the pellet was resuspended with 1ml LB medium. The resuspended culture and 50% (v/v) glycerol were mixed into 4:1 ratio and flash frozen in liquid nitrogen and stored at -80°C.

2.8 Protein heterologous expression and protein purification

2.8.1 Expression of proteins in *Escherichia coli*

Target proteins were expressed from pET-32a recombinants by the induction of T7 promoter. The induction of expression was achieved by adding 1M IPTG to a final concentration 0.4 mM in the cell culture Origami^{TM} 2 (DE3) containing recombinant expression vector. The process was carried out in LB medium with 50 μg ml^{-1} carbenicillin and streptomycin.

The cell cultures were collected after the required expression time. The cultures were stored at -20°C or used for protein extraction immediately.

The bacterial cells were washed by 1 x PBS buffer (pH7.4) twice and resuspended in the buffer required by the protein purification material. Cells were briefly broken by a freeze/thaw treatment three times and were thoroughly broken by sonication. Part of the resultant extracts was collected as total cell fraction samples for verification. The remaining mixture was centrifuged at 20,000***g*** for 10 min to separate the soluble fraction and insoluble fraction for further analyses. SDS-PAGE and western blot were carried out to allow protein sample verification.

2.8.2 Dialysis

The MWCO (molecular weight cut-off) of dialysis tubing used in this study was 1.2 kD. The dialysis tubing was boiled in 2% sodium bicarbonate/1 mM EDTA for 10 min, following 10 min boiling in water. The tubing was rinsed with dialysate buffer (Appendix 2, 2.2.7) before use. The treated tubing was stored in 50% ethanol/1 mM EDTA at 4°C. Each sample was loaded into the tubing, then sealed by clips. Proteins were dialysed for 1 to 2 hours at room temperature and then dialysed 1 to 2 hours in fresh buffer. The dialysate buffer was then changed and dialysis continued overnight at 4°C. The dialysed protein samples were stored at -20°C or used for cleavage reaction (see section 2.8.3) immediately.

2.8.3 Enzyme cleavage of protein tags

Enterokinase light chain (New England Biolab) was used to cleave fused 'tags' from the target protein. The cleavage site 'Asp-Asp-Asp-Asp-Lys' can be specifically recognised by enterokinase. The reaction was carried out based on manufacturers provided protocol. The buffer of purified protein products was adjusted to buffer that is suitable for enterokinase by dialysis (see section 2.8.2, Appendix 2, 2.2.7).

The amount of required enterokinase varies from batch to batch. Thus, different quality ratios (0.0001-0.5%, w/w) of enterokinase to target protein were tested for the optimisation of the reaction.

2.8.4 Protein purification

The isolation of proteins with His-tags was carried out by His-Select® HF Nickel Affinity Gel as a mini trial (Sigma-Aldrich). The large scale of protein purification was carried out with HisTrap™ FF column (GE Healthcare).

The isolation of target protein from its cleaved tags was carried out with reverse phase chromatography by using a C18 column (Sep-Pak® Cartridge, Waters). The buffer system was established by using two types of buffer: buffer A (0.1% TFA in degassed deionised water) and buffer B (0.1% TFA in acetonitrile). Before loading the protein sample, the column was equilibrated by 10ml of 5% buffer B. The protein sample was diluted 10 times with 50% buffer B before loading. After the protein sample was loaded, the column was

washed by at least 20ml of 5% buffer B. Different fractions of the protein sample were eluted by 2ml of 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 80% buffer B, and collected into 1.5ml tubes. The flow rate of the column was maintained as 0.375ml/min. Each eluted product was dried in freeze drier and dissolved in 200mM PBS pH 7.4 for further analyses.

2.9 Protein verification and proteomic analysis

2.9.1 SDS-PAGE

The gel solution preparing and sample detecting protocol can be found in Appendix 2.2.3. Each sample was diluted by using same volume of 2x sample buffer (Appendix 2.2.3). The samples were heated at 95°C for 10min before loading. The electrophoresis was carried out with MiniPROTEAN® 3 Cell (Bio-Rad).

2.9.2 Protein gel staining

Proteins on SDS-PAGE gel were visualized by using Coomassie Brilliant Blue R-250 staining system (Appendix 2.2.3) or SilverXpress® Silver Staining Kit (Thermo Fisher Scientific).

2.9.3 Western blotting

The protein samples were separated by SDS-PAGE. The gel was equilibrated in transfer buffer. The polyvinylidene difluoride (PVDF) membrane was soaked in methanol, H_2O and transfer buffer. The membrane, filter paper and fibre pads were soaked in the transfer buffer for 15 minutes. The transfer cassette (Figure 2.1) and frozen Bio-ice cooling unit was placed into the module. The tank was filled with cooled transferred buffer. Transfer process was carried out with MiniPROTEAN® 3 Cell (Bio-Rad) at 100V for one hour. After removing the membrane, the membrane was dried on filter paper and wetted in methanol. The membrane was then rinsed in H_2O or two minutes and blocked by 50 ml blocking buffer for one hour. The membrane was incubated with 5 ml 5000 times diluted primary antibody (6x His Rabbit Polyclonal antibody, Genetex) in a sealed plastic bag

overnight at 4°C with shaking. The membrane was washed with 50 ml blocking buffer for 10 minutes three times. The membrane was then incubated in 30 ml 30,000 x diluted secondary antibody (goat anti-rabbit IgG-HRP, Invitrogen) for at least 1.5 hours. After the membrane was washed with blocking buffer (as last washing step), the membrane was dried and proteins combined with antibodies were detected by using Amersham ECL Advance Western Blotting Detection Kit (GE Healthcare) X-ray film (Fuji).

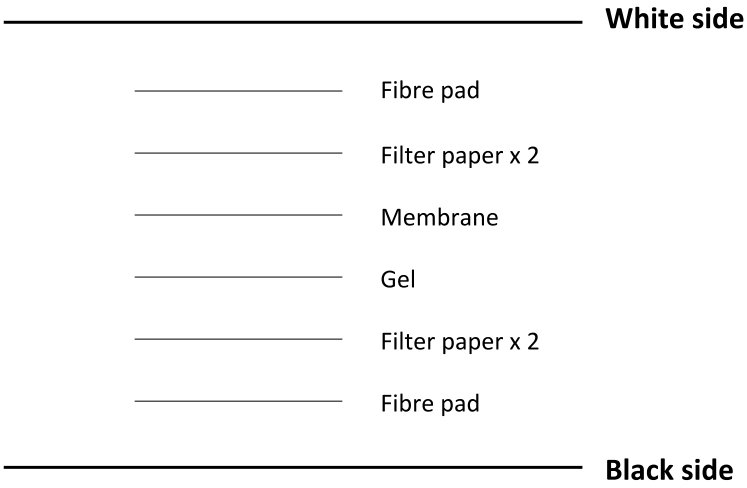


Figure 2.1 | Diagram of transfer cassette.

2.9.4 Mass spectrometry

The mass spectrometry carried out in this study was done by Dr Kate J Heesom at Proteomic Facility at the University of Bristol. The Samples in 20 mM Tris-HCl pH 8 were reduced by using 10mM TCEP and incubated for 1 hour at 55°C, alkylated by using 40mM iodoacetamide and incubated for 1 hour at room temperature, following an incubation with 2.5% (w/w) trypsin overnight at 37°C. Following digestion, the samples were resuspended in 5% formic acid and desalted using Sep-Pak® cartridges according to the manufacturer's instructions (Waters, Milford, Massachusetts, USA). Eluate from the Sep-Pak® cartridge was evaporated to dryness and resuspended in 1% formic acid prior to LC-MS/MS. Samples were fractionated by using an Ultimate 3000 nano-HPLC system in line with an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Peptides in 1% (v/v) formic acid were injected onto an Acclaim PepMap™ C18 Nano-Trap column (Thermo Fisher Scientific). After washing with 0.5% (v/v) acetonitrile 0.1% (v/v) formic acid peptides were resolved on a 250 mm × 75 μm Acclaim PepMap™ C18 reverse phase analytical column (Thermo Fisher Scientific) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1 min, 6-15%B over 58 min, 15-32%B over 58

min, 32-40%B over 5 min, 40-90%B over 1 min, held at 90%B for 6min and then reduced to 1%B over 1 min.) with a flow rate of 300 µl min^{-1}. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.1 kV using a stainless steel emitter with an internal diameter of 30 µm (Thermo Fisher Scientific) at a capillary temperature of 250°C. Tandem mass spectra were acquired using an LTQ-Orbitrap Velos mass spectrometer controlled by Xcalibur 2.1 software (Thermo Fisher Scientific) and operated in data-dependent acquisition mode. The Orbitrap was set to analyse the survey scans at 60,000 resolution (at m/z 400) in the mass range m/z 300 to 2000 and the top twenty multiply charged ions in each duty cycle selected for MS/MS in the LTQ linear ion trap. Charge state filtering, where unassigned precursor ions were not selected for fragmentation, and dynamic exclusion (repeat count, 1; repeat duration, 30s; exclusion list size, 500) was used. Fragmentation conditions in the LTQ were as follows: normalized collision energy, 40%; activation q, 0.25; activation time 10ms; and minimum ion selection intensity, 500 counts.

The raw data files were processed and quantified using Proteome Discoverer software v1.4 (Thermo Fisher Scientific) and searched against the UniProt *Arabidopsis thaliana* (50990 entries) or *Brassica oleracea* (58555 entries) databases, as well as the reverse decoy database (the same database but with all the protein sequences reversed) by using the SEQUEST algorithm. Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.8 Da. Search criteria included carbamidomethylation of cysteine (+57.0214) as a fixed modification and oxidation of methionine (+15.9949) as a variable modification. Searches were performed with full tryptic digestion and a maximum of 1 missed cleavage was allowed. Each match of a spectrum to a peptide was given a score based on how closely the spectrum matches the predicted given peptide sequences. Therefore, any matches to the decoy database are expected to have low scores. The match between the spectrum and the highest-scoring peptide is defined as a peptide-spectrum match (PSM). The PSMs were statistically validated to avoid false positives by using the False discovery rate (FDR)-controlling procedure. In this procedure, the distributions of the scores matched to both the genuine and the reversed decoy database was cut-off at which there is only a 5% chance (FDR=5%) that a peptide matched to the reversed decoy database. All the peptides with a score distributed below this cut-off threshold (FDR>5%) were excluded from the final data set. Peptides with scores fall between 1%<FDR<5%

were defined as medium confidence peptides, while peptides with scores fall at FDR<1% were defined as high confidence peptides.

2.9.5 Bradford assay

The protein concentrations of protein solution samples were measured by Bradford protein assay. The reaction was set up by mixing 1ml Bradford Reagent and 20µl protein sample. The reaction was incubated at room temperature for 5 min. The absorbance of reaction at 595nm was measured by using spectrophotometer (Spectroquant® Pharo 300). A standard curve was produced by using BSA (bovine serum albumin) as standard protein (Appendix 2, Figure S4.2). The linear concentration range was 0.1-1.4mg/ml. The protein sample concentrations were evaluated based on the linearisation equation:

y (A_{595})=0.3642x (concentration mg/ml)+0.4006

x (concentration mg/ml)= [y (A_{595})-0.4006]/0.3642

2.10 Isolation of proteins from pollen coat and stigmas

2.10.1 Pollen and stigma collection

200 mg Col-0 dry seeds was soaked in 1 L 0.1% agar for 48 hours at 4°C. Soaked seeds were planted on 40 trays of pesticide Intercept systemic insecticide treated F2+S compost. The plants were grown in the GM glass house for 3-4 weeks until flowering. Pollen was collected daily by a vacuum cleaner with a filter system constructed from plumbing parts with two layers of interspersed mesh filters. Unwanted plant tissue collected on the 150-micron mesh filter, while pollen grains collected on the 10-micron mesh filter.

2.10.2 Protein isolation from pollen coat

A filter unit was prepared before extraction: at the bottom of a 0.5 ml tube, punctured a hole by using a 21G x 1 1/2 gauge needle. A 14 mm diameter disc of Glass Microfiber (Whatman®) was compacted at the bottom of the tube and the tube was placed into a clean 1.5 ml tube for collecting the eluate.

100 mg pollen grains were mixed with 1 ml cyclohexane and vortexed briefly, then separated the mixture evenly into four filter units. The units were centrifuged at ~16,000***g*** for 14 s to elute down the liquid phase. This was repeated until no more liquid appeared in the collecting tube. Freeze-dried the cyclohexane and left a honey-like precipitation. 100µl MilliQ water was added and chilled on ice. The mix was sonicated on ice until milky homogenous suspension was observed. The mix was centrifuged and a layer of lipid-like phase could be observed on the top of an aqueous phase. The aqueous phase was pipetted into another clean tube. This centrifuge process was repeated until the product is clear and no lipid-like particles could be observed. The pollen coat complement sample with protease inhibitor (cOmplete^{TM} Protease Inhibitor Cocktail Tablets, Sigma-Aldrich) was stored at -80°C for further analyses.

2.10.3 Protein isolation from stigmas

Stigmatic microsomal membrane fraction extraction method was modified from protocol described in (Doughty *et al.,* 1998). Approximately 300 stigmas were homogenised in liquid nitrogen with a micro-pestle and mixed with 100 µl of 50 mM potassium phosphate pH 7.5, 1% Triton X-114. The homogenate was centrifuged (5804R Eppendorf) at 4,000***g*** at 4°C for 10 minutes to remove the cellular debris. This centrifuging was repeated once, the supernatant was then transfered to a polycarbonate centrifuge tube. The supernatant was ultracentrifuged (L-80 Beckman Coulter) at 100,000***g*** at 2°C for 30 minutes. The supernatant was collected as total cell protein (TCP) solution and the pellet was washed with 200 µl of 50mM potassium phosphate pH 7.5. The ultracentrifuge step was repeated once. The pellet was solubilised for 30 minutes on ice with 100 µl 50 mM potassium phosphate pH 7.5, 1% Triton X-114. The solubilised mixture was ultracentrifuged at 100,000***g*** at 2°C for 30 minutes to remove any insolubilised fraction. The supernatant was solubilised membrane protein solution. Protease inhibitor was added to each extraction product that was stored at -20°C for further analyses.

2.11 Protein-protein interaction

2.11.1 Far-western blotting

The total cell protein sample was separated by SDS-PAGE and transferred onto PVDF membrane (see section 2.9.3), 15 μg of purified fusion AtPCP-Bγ in 7.25 ml of blocking buffered was sealed with the membrane in a plastic bag and incubated overnight at 4°C. Standard western blotting was carried out after the blotting membrane was washed with 50ml blocking buffer for 10 minutes three times.

2.11.2 Protein pull-down assay

His-Select® HF Nickel Affinity Gel (Sigma-Aldrich) was used for protein pull-down assay of AtPCP-Bγ and stigmatic membrane proteins. AtPCP-Bγ and stigmatic membrane protein solution was prepared as described in secsion 2.8 and 2.10, respectively. All steps were carried out at room temperature. 50μl of His-Select® HF Nickel Affinity Gel was added into each 1.5ml microcentrifuge tube and centrifuged (5415D, Eppendorf) for 30 seconds at 5,000***g*** to remove ethanol. Before loading the samples, the gel was washed by 100 μl of MilliQ water once and 200 μl of binding buffer (Appendix 2.2.6) twice. 200 μl of AtPCP-Bγ solution (in binding buffer, Appendix 2.2.6) was added into gel and mixed gently for 5 minutes. The same volume of binding buffer was added into another tube with 50 μl washed gel as a control. The samples were centrifuged at 5,000***g*** for 30 seconds to remove supernatant. The gel was washed with 500 μl of binding buffer for three times to remove any unbound proteins. 20 μl of stigmatic membrane protein solution was added into the gel binding with AtPCP-Bγ and the control gel without AtPCP-Bγ. The mixtures were gently mixed for 20 minutes and centrifuged at 5,000***g*** for 30 seconds. The supernatant was discarded and the gel was washed with 500 μl of binding buffer three times. The binding fraction was then eluted with 100 μl of elution buffer (Appendix 2.2.6). The eluted products were collected and stored at -20°C for further analyses, including spectrophotometry (MS).

2.11.3 Protein crosslinking

Glutaraldehyde (GTA) was used as crosslinker in this protein crosslinking assay. Thus, all the buffers were free from amines. 2.5% GTA was prepared by diluting 25% of GTA

solution 10 times. 50 µl of AtPCP-Bγ solution and 50µl of stigmatic membrane protein solution was mixed together in 50mM phosphate buffer pH 7.5. 5 µl of 2.5% GTA was added to the mixture and gently mixed well. The mixture was incubated at 25°C for 15 minutes. The AtPCP-Bγ solution was replaced with 50 µl of 50 mM phosphate buffer pH 7.5 in another tube as a control. The reaction was quenched by adding 2 µl of 1M Tris-HCl pH 8.0 at the end of incubation. The products were purified by using His-Select® HF Nickel Affinity Gel to isolate the crosslinked fractions with His-tag. The purified fraction was sent to mass spectrophotometry (MS) (see section 2.9.4) for analysis of the crosslinked proteins.

Chapter 3 Identification and functional characterisation of AtPCP-Bs

The content of this chapter has been published in New Phytologist (Accept date: 23rd July, 2016). The final revised manuscript is shown in 3.2 (with supporting information in Appendix 1). The online version is attached at the end of this thesis.

Contributions:

James Doughty and Ludi Wang planned and designed the research and wrote the manuscript. Ludi Wang was involved with all aspects of the research including the phylogenetic analysis of PCP-Bs and homologues, protein sequence alignment, genotyping of T-DNA lines, production of double and triple mutant lines, gene expression analyses by using RT-PCR and GUS staining, pollen hydration, adhesion, pollen tube growth assays and data analyses, SEM and TEM work, prediction and modelling of PCP-B protein structures. Lisa A Clarke and Russell J Eason contributed to the expression analysis by using *in situ* hybridisation. Christopher C Parker contributed to part of the SEM work. James Doughty carried out the RNA gel blot analysis, protein purification, N-terminal sequencing and RACE PCR cloning of *Bo*PCP-B1 and *Bo*PCP-B2 during previous projects.

3.1 Introduction

The early stages of post-pollination in angiosperms involve multiple phases of interactions between male and female reproductive tissues (Chapter 1, 1.2, Figure 1.5). The establishment of pollen-stigma compatibility is a key early step in plant reproduction, which enables compatible pollen to be recognised by the receptive papillar cells. Despite the fundamental importance of pollen-stigma compatibility very little is known about the molecular regulation of this recognition system. Previous work has revealed that the pollen coat of members of the Brassicaceae carry a diverse array of proteins with most having unknown functions (unpublished observations) (Ross & Murphy, 1996; Mayfield *et al.*, 2001). Importantly, many pollen coat proteins fall into a broad family of small secreted cysteine-rich proteins (CRPs) which have often been shown to act as ligands to protein receptor targets at the plasma membrane (see Chapter1 1.3). Of plant CRPs a number have been identified as having various roles in plant reproduction (reviewed in Marshall *et al.*, 2011) (see section 1.3.2). Of particular relevance here is the fact that CRPs found in the

pollen coat of members of the Brassicaceae, the pollen coat proteins (PCPs), are emerging as important regulators of the pollen-stigma interaction (Chapter 1 1.4) (Doughty *et al.*, 1998; Schopfer *et al.*, 1999; Takayama *et al.*, 2000). One class of PCPs isolated from the pollen coat of *Brassica oleracea*, the PCP-Bs, have previously been described, but their function was unknown (Doughty *et al.*, 2000).

Preliminary work carried out in the lab showed the expression of four PCP-Bs in Arabidopsis pollen. Possible regulatory function of PCP-Bs in the pollen hydration and adhesion were demonstrated by a range of preliminary bioassays. We hypothesised that members of this group of proteins are likely to be key mediators of pollen compatibility acting at the earliest stages of the pollen-stigma interaction.

In this project, four putative *Arabidopsis thaliana* PCP-B-encoding genes were identified, determined to be gametophytically expressed during the late stages of pollen development and confirmed as pollen coat proteins. Modified from previous methods, *in vivo* bioassays utilising single and multiple *pcp-b* gene knockouts revealed that *At*PCP-Bs function in the early stages of pollination. Pollen morphology was unaffected in *pcp-b* lines, however mutant pollen grains showed striking defects in pollen hydration, delays in pollen tube emergence, as well as weakened anchoring of pollen grains to the stigma surface. This evidence suggests that *At*PCP-Bs are important components of the basal compatibility system, establishing a molecular dialogue between compatible pollen grains and the stigma.

3.2 Manuscript

PCP-B class pollen coat proteins are key regulators of the hydration checkpoint in *Arabidopsis thaliana* pollen-stigma interactions

Ludi Wang[1], Lisa A Clarke[1], Russell J Eason[1], Christopher C Parker[1], Baoxiu Qi[1], Rod J Scott[1], James Doughty[1]

[1]Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK

Summary

- The establishment of pollen-pistil compatibility is strictly regulated by factors derived from both male and female reproductive structures. Highly diverse small cysteine-rich proteins (CRPs) have been found to play multiple roles in plant reproduction, including the earliest stages of the pollen-stigma interaction. Secreted CRPs found in the pollen coat of members of the Brassicaceae, the pollen coat proteins (PCPs), are emerging as important signalling molecules that regulate the pollen-stigma interaction.
- Using a combination of protein characterisation, expression, and phylogenetic analyses we identified a novel class of *Arabidopsis thaliana* pollen-borne CRPs, the PCP-Bs (for pollen coat protein B-class) that are related to embryo surrounding factor (ESF1) developmental regulators. Single and multiple *PCP-B* mutant lines were utilised in bioassays to assess effects on pollen hydration, adhesion and pollen tube growth.
- Our results reveal that pollen hydration is severely impaired when multiple PCP-Bs are lost from the pollen coat. The hydration defect also resulted in reduced pollen adhesion and delayed pollen tube growth in all mutants studied.
- These results demonstrate that *At*PCP-Bs are key regulators of the hydration 'checkpoint' in establishment of pollen-stigma compatibility. In addition we propose that interspecies diversity of PCP-Bs may contribute to reproductive barriers in the Brassicaceae.

Key words: pollen-stigma interaction, compatibility, pollen coat proteins, pollen hydration, signalling, reproduction, *Arabidopsis thaliana*

Introduction

Pollination in angiosperms involves multiple phases of interaction between female reproductive tissues of the gynoecium and the male reproductive unit, pollen, and subsequently the pollen tube it produces on germination (Hiscock & Allen, 2008). The process is highly selective such that the majority of heterospecific pollen fails to effect syngamy and indeed intraspecific pollination can also be blocked in those species which possess self-incompatibility (SI) systems. Such prezygotic reproductive barriers are evolutionarily advantageous as they limit wasted mating opportunities, contribute to reproductive isolation and facilitate outbreeding when SI is present (Yost & Kay, 2009; Smith *et al.*, 2013). The establishment of compatibility is complex and involves a suite of biophysical and molecular recognition factors that operate throughout the pollination process; from pollen capture by the stigma, pollen hydration, germination, stigmatic penetration to polarised tube growth through the pistil (reviewed in Chapman & Goring, 2010). Although much progress has been made in elucidating the mechanisms that regulate compatibility in a broad range of species, it is clear that great mechanistic diversity exists, with no common system in operation (Hiscock & Allen, 2008; Allen *et al.*, 2011). Such diversity is considered to be important in maintaining species barriers (Swanson & Vacquier, 2002; Takeuchi & Higashiyama, 2012).

Members of the Brassicaceae family (which includes *Brassica* and *Arabidopsis* species) are characterised by having stigmas of the 'dry' type, which lack sticky secretions such as those present in species of the Solanaceae (Heslop-Harrison & Shivanna, 1977). Dry stigmas are highly discriminatory, reducing the probability that heterospecific pollen grains and pathogenic spores will be captured, hydrate and germinate on their surfaces. In this system compatibility is established at the stigma surface within minutes of pollination with compatible grains gaining access to stigmatic water whereas incompatible pollen generally fails to fully hydrate and germinate. Thus, pollen hydration on the stigma surface is essential for successful reproduction and is a strictly regulated checkpoint centred in the stigma (Dickinson, 1995; Ma *et al*., 2012; Hiroi *et al*., 2013). Possession of a dry stigma comes with the requirement that the exine surface of conspecific pollen must carry a coating (tryphine) (Dickinson, 1995; Dickinson *et al.*, 2000). Tryphine is a complex mixture of lipids, proteins, glycoconjugates and pigments (Piffanelli *et al.*, 1998; Hernandez-Pinzon *et al.*, 1999) that confers adhesive properties to the grain, provides a

conduit for water to pass from the stigma to effect pollen hydration and, importantly, carries factors that determine compatibility (Dickinson, 1995; Safavian & Goring, 2013). Pollen access to stigmatic water requires targeted secretion in the stigma immediately adjacent to a compatible pollen grain (Dickinson, 1995) and it is now well established that in the Brassicaceae this involves exocyst-mediated tethering of secretory vesicles to the stigmatic plasma membrane (Samuel *et al.*, 2009; Safavian & Goring, 2013; Safavian *et al.*, 2015).

Despite progress in identifying molecular regulators of compatibility in stigmas relatively little is known about the pollen-borne signals that establish it. Components of the pollen coat are most likely to mediate compatibility due to the intimate interaction of this layer with the surface of stigmatic papilla cells and the speed of pollen acceptance (Elleman & Dickinson, 1986; Elleman & Dickinson, 1990; Preuss *et al.*, 1993). Indeed application of isolated pollen coat to the stigma surface evokes a rapid expansion of the stigmatic outer wall layer (Elleman & Dickinson, 1990; Elleman & Dickinson, 1996) and isolated pollen stimulates the production of structures resembling vesicles in the stigma apoplast beneath the pollen contact site (Elleman & Dickinson, 1996).

Analysis of pollen coat components and mutational studies in *A. thaliana* have shed light on factors that influence the pollen-stigma interaction. A number of these appear to be biophysical in nature, for example *eceriferum* (*cer*) mutant pollen fails to hydrate due to the elimination of very long chain lipids from the pollen coat (Preuss *et al.*, 1993; Hulskamp *et al.*, 1995; Fiebig *et al.*, 2000). Hydration defects have also been reported in mutants for extracellular lipase 4 (EXL4) and GRP17, an oleosin-domain-containing glycine-rich protein which may work cooperatively to alter the lipid composition at the pollen-stigma interface to facilitate the passage of water to the grain (Mayfield & Preuss, 2000; Updegraff *et al.*, 2009). Work in *Brassica* has led to the identification of several groups of small cysteine-rich pollen coat proteins (Doughty *et al.*, 1993; Hiscock *et al.*, 1995; Doughty *et al.*, 1998; Schopfer *et al.*, 1999; Doughty *et al.*, 2000; Takayama *et al.*, 2000; Shiba *et al.*, 2001). Importantly these polypeptides, rather than having major effects on biophysical properties of the pollen coat, act as ligands and have been demonstrated to bind a number of stigmatic proteins.

In recent years, a broad range of cysteine-rich proteins (CRPs) have been identified in plants having functions in cell signalling, development and defense (Silverstein *et al.*, 2007; Li *et al.*, 2014). They are all characterised by being small (less than 160 amino acids), having a conserved N-terminal signal peptide and a C-terminal cysteine-rich region with the pattern of cysteines determining their classification (Marshall *et al.*, 2011). A number function in plant reproduction, including pollen-stigma self-recognition (Schopfer *et al.*, 1999; Shiba *et al.*, 2001), pollen tube growth and guidance (Chae *et al.*, 2009; Okuda *et al.*, 2009) and early embryo development (Marshall *et al.*, 2011; Costa *et al.*, 2014). Several families of CRPs have been identified in the pollen coat of *Brassica* and *Arabidopsis,* and some have confirmed roles in the pollen-stigma interaction. In self-incompatible species, the *S*-locus cysteine-rich protein (SCR/SP11) (Schopfer *et al.*, 1999; Shiba *et al.*, 2001), acts as the male determinant and interacts with the *S*-receptor kinase (SRK) (Takasaki *et al.*, 2000), to trigger pollen rejection by targeted degradation of the basal compatibility factor EXO70A1 (Stone *et al.*, 2003; Samuel *et al.*, 2009). Other CRPs belonging to the PCP-A class of *Brassica* pollen coat proteins such as SLR-BP1 and PCP-A1, bind the stigmatic proteins *S*-locus related 1 (SLR1) and *S*-locus glycoprotein (SLG) respectively and thus likely function in the pollen-stigma interaction, though their precise function remains to be determined (Doughty *et al.*, 1998; Takayama *et al.*, 2000). A further class of *Brassica* pollen coat CRPs, the PCP-Bs, have been described and are also good candidates for regulators of the earliest phases of the pollen-stigma interaction (Doughty *et al.*, 2000). Thus there is a growing body of evidence demonstrating that the pollen coat carries factors that mediate both incompatibility and compatibility and that cysteine-rich pollen coat proteins are important to molecular dialogue in the pollen-stigma interaction. Although many studies have focused on the mechanisms of self-incompatibility (SI) in a range of species, the molecular regulation of self-compatibility (SC) in flowering plants is still poorly understood.

In this study, we report on the identification of four *Arabidopsis thaliana* PCP-B encoding genes termed *AtPCP-Bα* (*At5g61605*), *AtPCP-Bβ* (*At2g29790*), *AtPCP-Bγ* (*At2g16535*) and *AtPCP-Bδ* (*At2g16505*), which are expressed gametophytically late in pollen development. By utilising T-DNA insertion lines carrying single, double and triple *AtPCP-B* gene knockouts, we examined the impact of these mutations on pollen morphology and the pollen-stigma interaction. Phenotypic analyses revealed defects in pollen hydration and delays in pollen tube growth for single and combined mutants compared with wild-type.

Triple mutant *pcp-bα/β/γ* pollen displayed a substantially reduced hydration rate on stigmas, delayed pollen tube growth, as well as weakened anchoring to the stigma surface. Importantly, no impact on pollen morphology was revealed in this study though the mutants presented striking effects on early post-pollination events. Such evidence suggests the *At*PCP-Bs act as important regulatory factors during the earliest stages of the pollen-stigma interaction by establishing a molecular dialogue between the stigma and pollen grains.

Materials and Methods

Plant material and growth conditions

Brassica oleracea var *alboglabra* L. homozygous for incompatibility haplotypes S25 and S29 (Horticultural Research International, Wellesbourne, UK) was used for isolation of *Bo*PCP-B1 and *Bo*PCP-B2 pollen coat proteins respectively. *Arabidopsis thaliana* (L.) Heynh. T-DNA insertion lines SALK_207087, SALK_062825, SALK_072366 (Alonso *et al.*, 2003) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). GABI_718B04 was purchased from GABI-KAT (Kleinboelting *et al.*, 2012). T-DNA insertion sites and their respective mutant alleles are detailed in Fig. S1.1. Single gene T-DNA insertion lines were backcrossed to wild-type (Col-0) at least three times before phenotyping. *pcp-bβ/γ* and *pcp-bα/β/γ* mutant lines were created using standard crossing procedures. *PCP-B* transcript status for each mutant was confirmed by reverse transcription polymerase chain reaction (RT-PCR) (Fig. S1.2). Primers for RT-PCR are detailed in Table S1.1. The A9-barnase male sterile line was provided by Rod Scott, University of Bath, UK (Paul *et al.*, 1992). GUS (β-glucuronidase) reporter lines *pAt5g61605:GUS* and *pAt2g16505:GUS* were provided by José F. Gutierrez-Marcos (University of Warwick, UK).

A. thaliana plants were propagated in Levington F2+S compost (Soils HS Limited, Wotton-Under-Edge, UK) in a controlled environment room with a 16h:8h, light:dark photoperiod provided by fluorescent lighting (130 μmol $m^{-2}sec^{-1}$). Temperature was

maintained at 21±1°C with 60% relative humidity. Brassicas were grown in a glasshouse at 21°C with a 16h:8h, light:dark photoperiod.

Reverse transcription polymerase chain reaction and RNA gel blot analysis

Anthers were collected from *A. thaliana* stage 10-12 flower buds (Smyth *et al.*, 1990), stigmas from open flowers of the *A. thaliana* A9-barnase line, roots from two-week-old seedlings grown on ½ MS plates and leaves from fully-grown rosettes. RNA was extracted using a PureLink® RNA Mini Kit (Thermo Fisher Scientific). cDNA synthesis was carried out using the ProtoScript® II First Strand cDNA Synthesis Kit (New England Biolabs). DNA amplification utilised DreamTaq® Green PCR Master Mix (2X) (Thermo Fisher Scientific). Primers for DNA amplification are detailed in Table S1.1. RNA gel-blot analysis was carried out as described previously (Doughty *et al.*, 1998) using polyadenylated mRNA (450 ng) from leaves, stigmas, pollen, and anthers derived from a range of bud sizes (whole flower buds for anthers of size <2 mm). Labelling of gene specific *BoPCP-B* probes (covering the coding region of the gene from aa residue 40 to the C-terminus) was conducted using the Prime-a-Gene Labeling System (Promega), with modifications to the deoxynucleotidetriphosphates mix to permit double labelling with dATP and dCTP (α-^{32}P, 100 µCi, 6000 Ci mmol^{-1} each).

Pollen hydration assays

For *in vivo* hydration assays, pollen grains derived from Col-0, *pcp-bα* (SALK_207087), *pcp-bβ* (SALK_062825), *pcp-bγ* (SALK_072366), *pcp-bδ* (GABI_718B04), *pcp-bβ/γ* and *pcp-bα/β/γ* lines were applied to stigmas of the A9-barnase male sterile line. Freshly opened mature flowers were used (retained on the plant) and pollen was applied in a monolayer with an eyelash. At least eight independent stigmas were used in assays from each line. For *in vitro* hydration assays pollen from Col-0 and *pcp-bα/β/γ* triple knockout lines were placed on a slide in a humid chamber (100% relative humidity). Pollens on stigmas were photographed under a dissecting microscope immediately after pollinations were initiated (designated as time point zero). Subsequent images were captured every minute for 30 minutes. For humid chamber assays pollen was photographed one minute after pollen grains were placed, then the chamber was sealed, after which images were

taken every minute for 30min. Equatorial diameter of pollen was measured in pixels using ImageJ Software (Schneider *et al.*, 2012). Pollen hydration (%) was calculated using the equation: Pollen hydration (%) = (Pollen diameter – initial pollen diameter)/ initial pollen diameter. Slopes were determined using 11 data points during the 0-10min, 10-20min, or 20-30min time periods using the linear regression curve $f = a_0 x + b$. All statistical analyses were carried out using Microsoft Excel 2013.

Pollen adhesion assay

Stigmas of *A. thaliana* A9-barnase plants were hand-pollinated using freshly dehiscent anthers from Col-0 and *pcp-bα/β/γ* lines. Pollen was applied as a monolayer. After 30 minutes the flower was excised from the plant and placed into 0.5ml of fixative (60% v/v ethanol, 30% v/v chloroform, 10% v/v acetic acid) in a 1.5ml microfuge tube. The sample was immediately shaken 10 times using short sharp strokes to dislodge pollen that was not strongly adherent to the stigma. The flower was then removed and placed into a separate microfuge tube. Both samples were retained for pollen counting. 50µl of aceto-orcein stain (1%) was added to the tubes and incubated overnight at room temperature (RT). 'Washed-off' pollen samples were centrifuged at 13,000***g*** (10 minutes), excess fixative was removed and the pellet was resuspended in 10µl of 50% glycerol before counting. Stigmas were excised from stained flowers before mounting in 50% glycerol and were squashed on a slide to ensure all pollen was visible for counting.

Pollen tube growth assay

Pollinations were initiated on *A. thaliana* male sterile A9-barnase stigmas and allowed to proceed for two or four hours before stigmas were excised and incubated in fixative (60% v/v ethanol, 30% v/v chloroform, 10% v/v acetic acid) overnight. After removal of fixative, stigmas were incubated in 8M NaOH for 20 minutes then washed in dH_2O three times, each for five minutes. Samples were transferred to 0.1% decolourised aniline blue (0.1% w/v aniline blue in 0.1M K_3PO_4, pH 11) for 1 hour before imaging (Kho & Baer, 1968).

Microscopy

Imaging of pollen hydration on stigmas, hydration in a humid chamber and GUS histochemical staining (Method S1.1) of flowers and leaves was carried out using a Nikon

SMZ1500 dissection microscope coupled to Nikon Digital Sight DS-U1 camera. A Nikon Eclipse 90i epifluorescence microscope (10 X objective) with Nikon Digital Sight DS-U1 camera was used for imaging pollen tubes stained with aniline blue and for anthers stained for GUS activity.

Scanning and Transmission Electron Microscopy

Pollinated stigmas were dry-fixed using the method of Elleman and Dickinson (1986). Samples were washed with 0.1M Sodium Cacodylate Buffer pH 7.4 and prepared for SEM and TEM by a modified method of Villar *et al*. (1987) using 2.5% glutaraldehyde and low viscosity resin. Samples for SEM were gold coated and imaged using a JEOL JSM640LV Scanning Electron Microscope (JEOL Tokyo Japan). Samples for TEM were ultra-thin sectioned (100nm) on an Ultracut-E ultramicrotome (Leica UK) and imaged by JEOL JEM1200EXII Transmission Electron Microscope.

Multiple sequence alignment and phylogenetic analysis

PCP-B-like sequences were retrieved from available complete plant genomes completed to at least scaffold level using TBLASTN database searches (Phytozome, https://phytozome.jgi.doe.gov; Comparative Genomics, CoGe, https://genomevolution.org). Nucleotide sequence alignments of *PCP-B* homologous genes were generated using MUSCLE (codon) (Edgar, 2004). Codons of protein coding sequences were translated into amino acid sequences before the alignment was performed. Aligned amino acid sequences were then replaced by the original codons. Graphical output of protein sequence alignment was generated by Jalview using 'ClustalX' colour coding. Phylogenetic trees were built using the Maximum Likelihood statistical method in MEGA (version 6) (Tamura *et al.*, 2013). The initial tree was determined by neighbour-joining method (NJ). The phylogeny test was carried out using the bootstrap method (1000 replications). Phylogenetic trees were displayed using iTOL (Letunic & Bork, 2007).

Protein structure prediction and modelling

Protein sequence alignment of PCP-Bs and 2RU1 was carried out with T-COFFEE (Notredame *et al.*, 2000). PCP-B protein models were built using SWISS-MODEL (Arnold *et al.*, 2006; Guex *et al.*, 2009; Kiefer *et al.*, 2009; Biasini *et al.*, 2014) based on the

modified T-COFFEE alignment result. Three-dimensional cartoon models and electrostatic potential surface models were produced by PyMOL (version 1.7.4).

In situ hybridisation

Flower buds were excised from inflorescences and immediately fixed in fresh 4% paraformaldehyde for 16 h at 4ºC with an initial 10 minutes under low vacuum. Tissues were embedded in Paraplast Xtra (Sigma), sectioned (7 to 10 µm), and prepared for probing as described by Langdale (1994), except for the protease treatment where sections were incubated for 30 min at 37ºC in 50µg mL^{-1} proteinase K (Sigma). Both antisense and sense probes were synthesized using a SP6/T7 digoxigenin RNA labeling kit (Boehringer Mannheim), according to the manufacturer's instructions. Probes covered the protein coding sequence of the *AtPCP-Bβ* and *AtPCP-Bγ* cDNAs.

Protein purification and N-terminal sequencing

*Bo*PCP-B1 and *Bo*PCP-B2 proteins were purified from total pollen coat proteins by a combination of gel filtration, RP-HPLC and cation exchange chromatography following the protocol described by Doughty *et al*. (Doughty *et al.*, 1993; Doughty *et al.*, 1998). Both *Bo*PCP-B1 and *Bo*PCP-B2 co-purified with the previously characterised PCP-A1 polypeptide following C18 RP-HPLC and were separated by cation exchange chromatography (Doughty *et al.*, 1998). *Bo*PCP-B1 and *Bo*PCP-B2 were isolated from *S*25 and *S29* incompatibility lines of *B. oleracea* var *alboglabra* respectively. *Bo*PCP-B protein samples were prepared for N-terminal sequencing as described previously (Doughty *et al.*, 1998).

5' and 3' Rapid Amplification of cDNA End (RACE) Polymerase Chain Reaction Cloning of *BoPCP-B1 and BoPCP-B2*

Polyadenylated RNA was isolated from 100mg of anthers derived from 9-11mm flower buds of *B. oleracea* var *alboglabra* (homozygous for S25 and S29 incompatibility haplotypes) using a QuickPrep Micro mRNA purification kit (Pharmacia Biotech). Cloning of *BoPCP-B1* and *BoPCP-B2* cDNA sequences was carried out using a 5'/3' RACE kit (Boehringer Mannheim, Lewes, UK) following the manufacturer's instructions. 1µg of mRNA was subjected to first-strand cDNA synthesis using an oligo(dT) kit primer. First

round 3' RACE cloning of *BoPCP-B1* utilised a degenerate primer based on the peptide sequence AGNAAK[P/Q] which is common to both *Bo*PCP-B proteins (5'- GC-GGATCC-GCIGGIAA[C/T]GCIGCIAA[A/G]C-3', where I represents inosine) in conjunction with a kit anchor primer. This was followed by two rounds of PCR using degenerate nested primers (5'- GC-GGATCC-AA[A/G]CA[A/G]ACICCITG[C/T]CA[C/T]G-3' and 5'- GC-GGATCC-AA[A/G]CCIAA[C/T]CA[C/T]ACITG) based on the *Bo*PCP-B1 specific peptide sequences KQTPCHE and KPNHTC respectively. 5' RACE was conducted utilising sequence specific primers SP1 and SP2 (derived from 3'RACE) for cDNA synthesis and nested amplification of the 5' region of the *BoPCP-B1* cDNA (SP1 5'-GCTTGCCGCACCTACGCG-3' and SP2 5'- CAT GTAGCACATGTTTTGAGC-3'). For *BoPCP-B2*, first round 3'RACE was carried out as described for *BoPCP-B1* followed by one further round of PCR utilising a degenerate nested primer (5'-GC-GGATCC-ATGAA[C/T]TG[C/T]GA[C/T]ACICA[A/G] G) based on the BoPCP-B2 specific peptide sequence MNCDTQD. 5' RACE utilised sequence-specific primers SP1 and SP2 (derived from 3'RACE sequence) for cDNA synthesis and nested amplification of the 5' region of the BoPCP-B1 cDNA (SP1 5'-GGCTTCCCAGATTTAGTGAC-3' and SP2 5'-GTGACACAACAAGAACAACTGCG-3').

Results

The pollen coat of *Brassica* contains polymorphic PCP-B class cysteine-rich proteins

In a previous study that characterised the SLG-binding pollen coat protein PCP-A1 from *Brassica oleracea* (Doughty *et al.*, 1998) two polypeptides were found to copurify with PCP-A1. These were purified to homogeneity and subjected to N-terminal sequencing. Each shared an identical six amino acid N-terminal domain and several conserved cysteine residues arranged in a unique pattern with respect to other known pollen coat protein families (Fig. S1.3). These were subsequently named *Bo*PCP-B1 and *Bo*PCP-B2 (for *B. oleracea* pollen coat protein, class B, 1 and 2 respectively). The partial *Bo*PCP-B polypeptide sequences permitted cloning of their respective full-length cDNAs by RACE PCR (GenBank accession numbers: *PCP-B1*, KX099662; *PCP-B2*, KX099663). *BoPCP-B1* and *BoPCP-B2* are predicted to encode proteins of 79 and 84 amino acids respectively with both having a putative 25 amino acid secretory signal peptide (Petersen *et al.*, 2011)

and a conserved pattern of eight cysteine residues in the mature protein. Based on the N-terminal sequence data, mature *Bo*PCP-B1 and *Bo*PCP-B2 are estimated to have M_rs of 5490 and 6109 respectively. The localisation of the *Bo*PCP-Bs to the pollen coat, together with their broad similarity to other small cysteine-rich proteins such as PCP-A1 (Doughty *et al.*, 1998) and the pollen self-incompatibility determinant SCR (Shiba *et al.*, 2001), suggested that they could potentially function in the pollen-stigma interaction.

PCP-Bs are evolutionarily widespread and have homology to Arabidopsis Embryo Surrounding Factor 1 (ESF1) developmental regulators

In order to facilitate subsequent functional analyses of PCP-B class pollen coat proteins putative homologues were identified in the model plant Arabidopsis. BLAST searching of the Arabidopsis genome using *Bo*PCP-B sequences revealed the presence of twelve *PCP-B*-like genes. All sequences were predicted to encode small, typically basic secreted proteins that shared a common cysteine pattern of seven or eight cysteines in the mature polypeptide (Fig. 3.1). Importantly, three members of this Arabidopsis gene family (*At1g10747*, *At1g10745* and *At1g10717*) encode the central cell-derived Embryo Surrounding Factor 1 (ESF1) signalling proteins known to shape early embryo development and patterning (Costa *et al.*, 2014). A phylogenetic analysis of mature protein-encoding gene regions, based on prior alignment of amino acid sequences, revealed that the *AtPCP-B*-like sequences fall in to two distinct clades (Fig. 3.2). One clade includes ESF1-encoding genes clustering into a group of five sequences. Of the other clade four of the sequences fall into a cluster which, following wider phylogenetic analysis across the *Brassica* and *Arabidopsis* genera, placed them in a clade that included genes encoding pollen coat-derived *Bo*PCP-Bs (Fig. 3.3). Expression analyses confirmed these four genes as being largely anther specific (Fig. 3.4). Taken together this data suggests that these Arabidopsis sequences are likely orthologues of the *B. oleracea PCP-Bs* and hence we named them *AtPCP-Bα* (*At5g61605*), *AtPCP-Bβ* (*At2g29790*), *AtPCP-Bγ* (*At2g16535*) and *AtPCP-Bδ* (*At2g16505*).

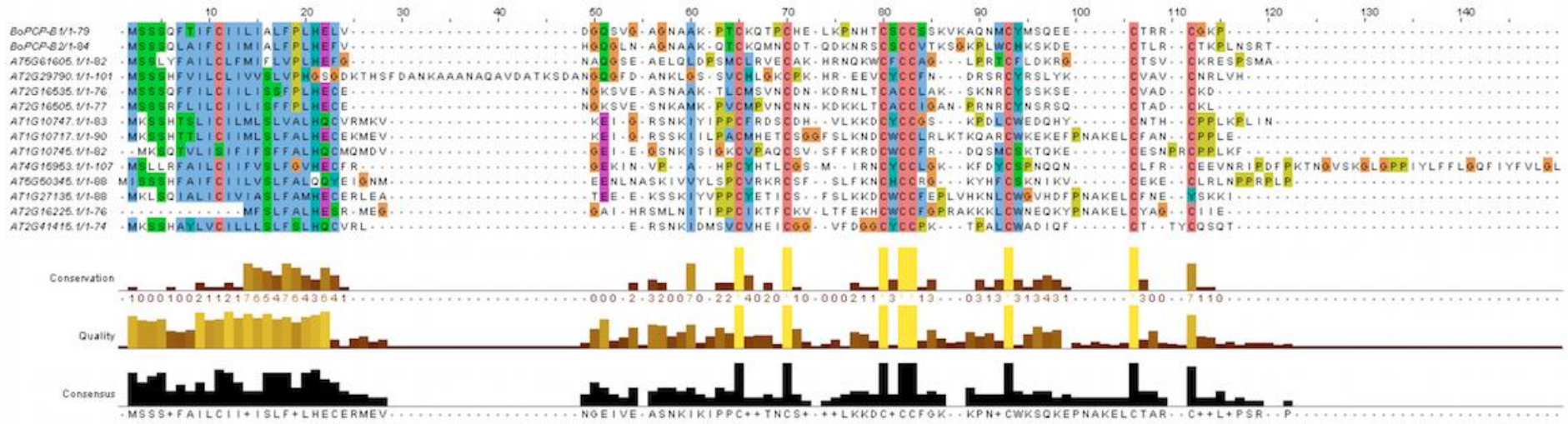


Fig. 3.1 Protein sequence alignment of *Brassica oleracea* PCP-B1 and PCP-B2 with all known *Arabidopsis thaliana* (Col-0) PCP-B-like proteins. *At*PCP-Bs are: *At*PCP-Bα (At5g61605), *At*PCP-Bβ (At2g29790), *At*PCP-Bγ (At2g16535) and *At*PCP-Bδ (At2g16505). At1g10747, At1g10745 and At1g10717 are ESF1.1, ESF1.2 and ESF1.3 respectively. Sequence conservation, quality and consensus is displayed below. Colour coding follows the default output for Clustal X (http://www.jalview.org/help/html/colourSchemes/clustal.html).

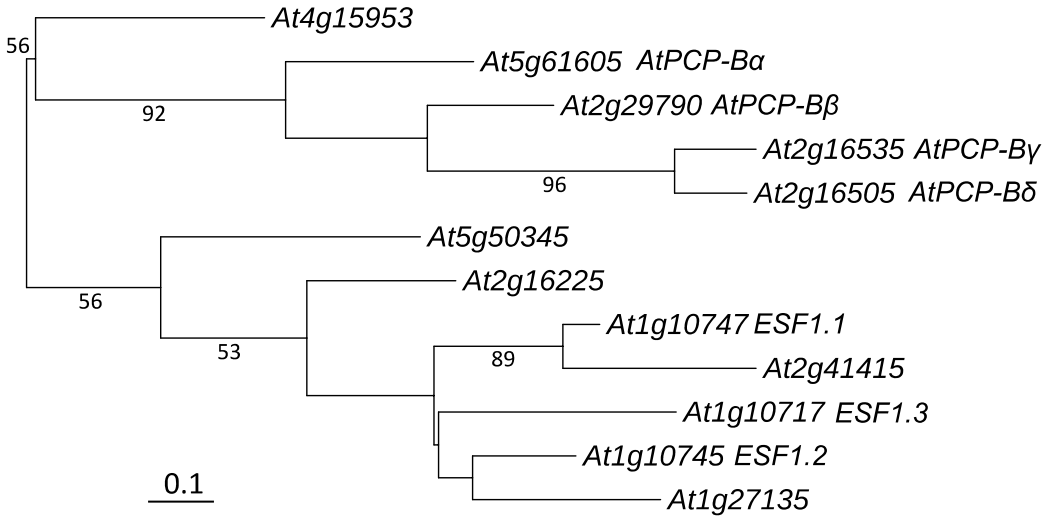


Fig. 3.2 Phylogenetic analysis of 12 *PCP-B* class genes in *Arabidopsis thaliana* (Col-0). The Maximum Likelihood tree was constructed by using the nucleotide sequences of predicted mature protein coding regions. Branch lengths are proportional to the scale bar, defined as 0.1 nucleotide substitutions per codon. The percentage bootstrap values (1000 re-samplings) higher than 50% are shown by interior branches.

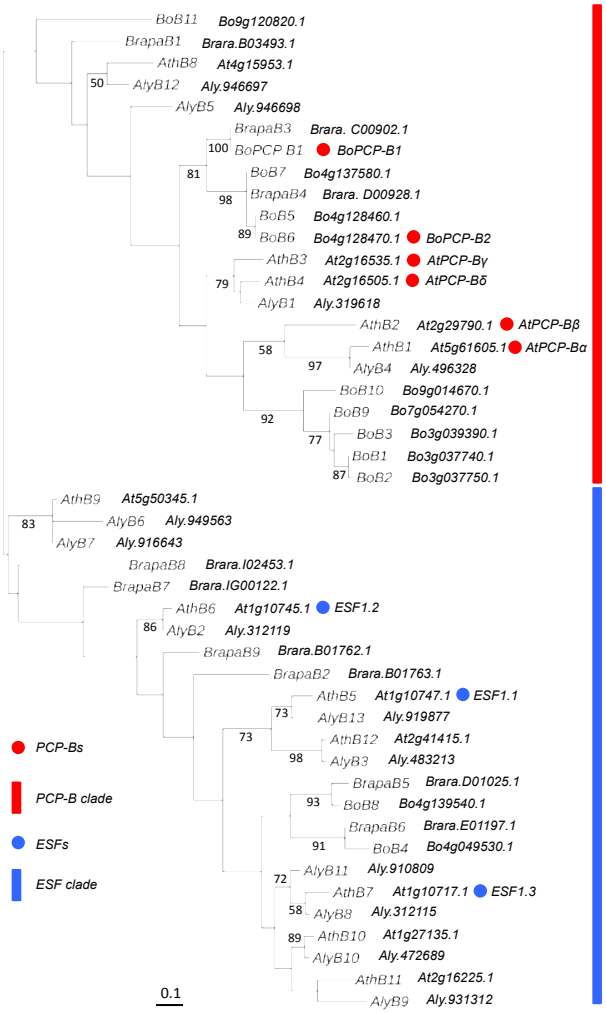


Fig. 3.3 Phylogenetic analysis of 46 *PCP-B* class genes in *Arabidopsis* and *Brassica*. The Maximum Likelihood tree was constructed using nucleotide sequences of the predicted mature protein coding regions. Bootstrap values (1000 re-samplings) higher than 50% are shown for interior branches. Branch length is scaled to the scale bar defined as 0.1 nucleotide substitutions per codon. The clades indicated by red and blue bars include *PCP-Bs* and *ESFs* respectively. Genes are abbreviated as: *AthB, Arabidopsis thaliana PCP-B-like; AlyB, Arabidopsis lyrata PCP-B-like; BoB, Brassica oleracea PCP-B-like; BrapaB, Brassica rapa PCP-B-like.* Gene loci or scaffolds are shown adjacent to gene abbreviations.

Iterative BLAST searches *Arabidopsis* and *Brassica* genera identified 46 PCP-B-like sequences in total across four species. Phylogenetic analysis of these protein sequences not only revealed the high degree of polymorphism across the family but also the presence of two distinct groupings of sequences, the *PCP-B* and *ESF1*-containing clades (Fig. 3.3). These groupings may reflect functional specialisation into seed and pollination-specific roles. Wider phylogenetic BLAST analyses across all known plant lineages identified 282 predicted PCP-B-like protein sequences in seven angiosperm families (36 species in total, Fig. S1.4, Table S1.2). In addition to the Brassicaceae, PCP-B-like proteins were found in

the Poaceae, Nelumbonaceae, Solanaceae, Malvaceae, Phrymaceae and Pedaliaceae. Thus the PCP-Bs are members of a wider family of highly polymorphic, though structurally related proteins, having an ancient evolutionary origin that predates the split between monocot and eudicot lineages.

Arabidopsis and *Brassica oleracea PCP-B* genes are expressed in maturing pollen

RNA gel-blot analysis for the *Brassica oleracea PCP-B1* and *PCP-B2* genes indicated high levels of expression in anthers (Fig. 3.4a). Transcripts were first detected at low levels in anthers derived from 5-7mm flower buds reaching a maximum by the 9-11mm bud stage by which time pollen is tricellular and tapetal cells that line the anther locule are fully degraded (Doughty *et al.*, 1998). This late pattern of expression in anther development infers that *Bo*PCP-Bs are likely to be gametophytically derived rather than being products of the tapetum. No expression was detected in leaves and stigmas, and only very low transcript levels were detected in mature pollen. This expression pattern exactly mirrors that of the pollen coat protein gene *PCP-A1* (Doughty *et al.*, 1998). In order to determine which of the twelve Arabidopsis *PCP-B*-like genes were likely orthologues of the *B. oleracea PCP-B* sequences expression analysis was carried out by RT-PCR (Fig. 3.4b). Six of the genes were found to be expressed in stage 12 anthers though two of these, At1g10747 and At1g10745, have previously been characterised as central cell-derived Embryo Surrounding Factor 1 (ESF1) signalling proteins involved in embryo patterning (Costa *et al.*, 2014). The remaining four anther-expressed genes (*At5g61605, At2g29790, At2g16535* and *At2g16505*, *AtPCP-Bα* to *δ* respectively) were found to share a similar temporal expression pattern to the *B. oleracea PCP-B*s (Fig. 3.4a, c). In addition, RNA-RNA *in situ* hybridisation for *AtPCP-Bβ* (Fig. 3.4d, S1.5) and *AtPCP-Bγ* and promoter-GUS fusions for *AtPCP-Bα* and *AtPCP-Bδ* (Fig. S1.6) confirmed high-level expression in pollen, further validating the status of this group as pollen coat protein-encoding genes. Taken together with the phylogenetic analysis that placed these four genes in the same clade as *BoPCP-B*s it is likely that *AtPCP-Bα*, *AtPCP-Bβ, AtPCP-Bγ* and *AtPCP-Bδ* are orthologous to the *BoPCPs*.

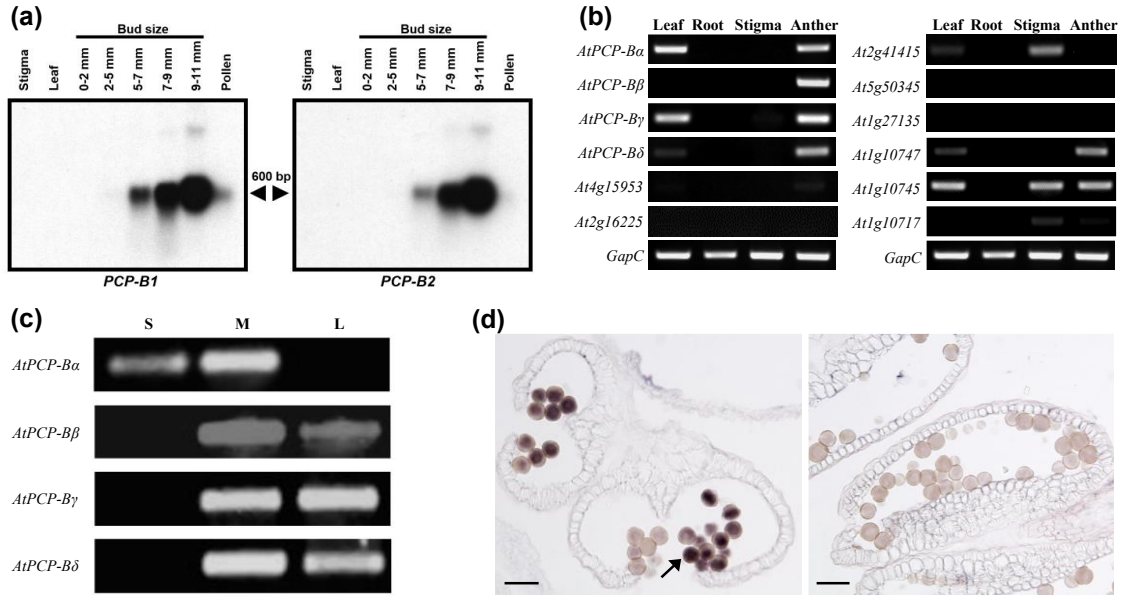


Fig. 3.4 Brassica and Arabidopsis *PCP-Bs* are gametophytically expressed late in pollen development. (a) mRNA gel blot analysis of *Brassica oleracea PCP-B1* and *PCP-B2* expression in leaves and reproductive tissue. Anthers from 9-11mm buds have a fully degenerated tapetum and pollen is trinucleate. The arrows indicate the size of the transcript in base pairs. (b) Reverse transcription polymerase chain reaction (RT-PCR) expression analysis of *AtPCP-B* and *AtPCP-B*-like genes in Arabidopsis leaves, roots, stigmas and anthers (derived from stage 12 buds). *GapC* - cDNA input control for RT-PCR. (c) *AtPCP-B* gene expression in flower buds through development (stages 10-12). S, small (<1mm; uninucleate microspores). M, medium (~1 to 1.5mm; binucleate pollen). L, large (>1.5mm; unopened buds, trinucleate mature pollen). Arabidopsis flower bud stages are as defined by Smyth *et al.* (1990). (d) RNA-RNA *in situ* hybridisation study of *AtPCP-Bβ* expression in *Arabidopsis thaliana* anthers. Image at left - transverse anther section treated with an antisense (+ve) *AtPCP-Bβ* DIG-labelled riboprobe, a clear signal (arrow) is observed within the majority of pollen grains. Image at right – longitudinal anther section treated with a control 'sense' (-ve) riboprobe with no signal being detectable in pollen grains. Scale bar: 20 μm.

Pollen hydration is impaired in *pcp-b* mutants

To investigate the effects of *AtPCP-B* gene mutations on early stages of the pollen-stigma interaction in *Arabidopsis thaliana*, *in vivo* pollen hydration assays were carried out by pollinating stigmas of the male sterile A9-barnase Col-0 line with pollen grains derived from wild-type and *pcp-b* plants. Four T-DNA insertion lines were identified as mutant alleles of *AtPCP-Bα, β, γ* and *δ* (Fig. S1.1) with *PCP-B* transcripts being undetectable in anthers for *pcp-bα-1*, *pcp-bβ-1* and *pcp-bγ*-1. PCP-*Bδ* expression was found to be substantially down-regulated, where the T-DNA insertion was located in the promoter region of the gene (Fig. S1.1, S1.2). No obvious vegetative or reproductive morphological abnormalities were observed in any of the lines. Each individual *pcp-b* line, a double

mutant (*pcp-bβ/γ)* and a triple mutant (*pcp-bα/β/γ*) were assessed utilising the pollen hydration assay. A quadruple mutant could not be generated due to the close genetic linkage of *PCP-Bγ* and *PCP-Bδ* (*c.* 9kb apart). Pollen equatorial diameter was recorded for 30 minutes following placement of pollen on stigmas. Four time points (0 min, 10 min, 20 min and 30 min) were selected for analysis of the difference of pollen hydration between wild-type and mutant lines. In addition, the rate of pollen hydration was assessed for each of the three 10 minute periods following pollination. On initiation of pollination (0 min) no significant difference was found between the diameters of pollen derived from wild-type and mutant lines (Fig. 3.5a). However at subsequent time points *pcp-bγ* and *pcp-bα/β/γ* pollen grains were significantly less hydrated than wild-type pollen (Fig. 3.5b-d). Assessment of pollen hydration as percentage change in diameter (from time point zero) demonstrated that the degree of pollen hydration was significantly lower in the *pcp-bβ, pcp-bγ*, *pcp-bβ/γ* and *pcp-bα/β/γ* mutant lines compared to wild-type at each time point (Fig. 3.5e-g). Despite there being no statistically significant difference in pollen hydration for *pcp-bα* and *pcp-bδ* compared to wild-type, median pollen diameters, hydration percentage and overall ranges in the data suggested *pcp-bα* and *pcp-bδ* mutations also negatively impact on pollen hydration (Fig. 3.5e-g).

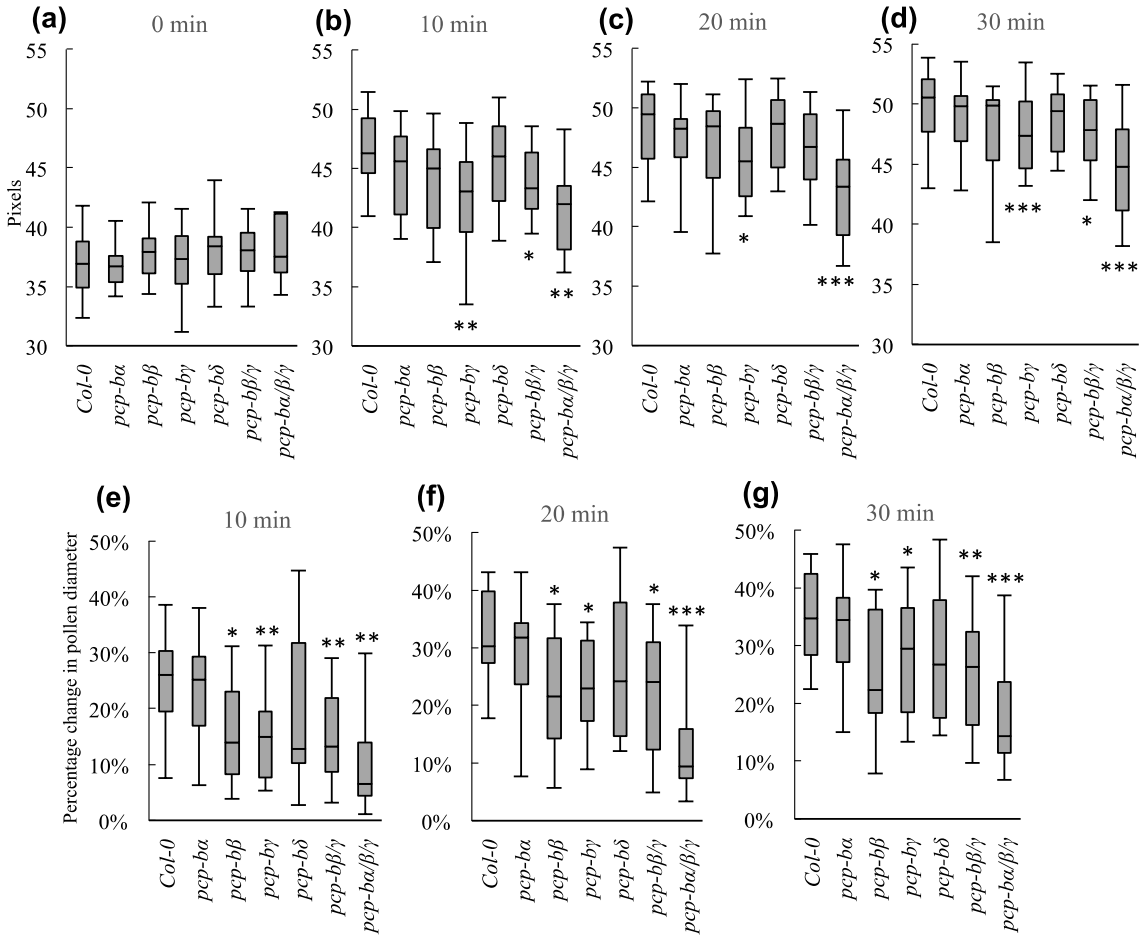


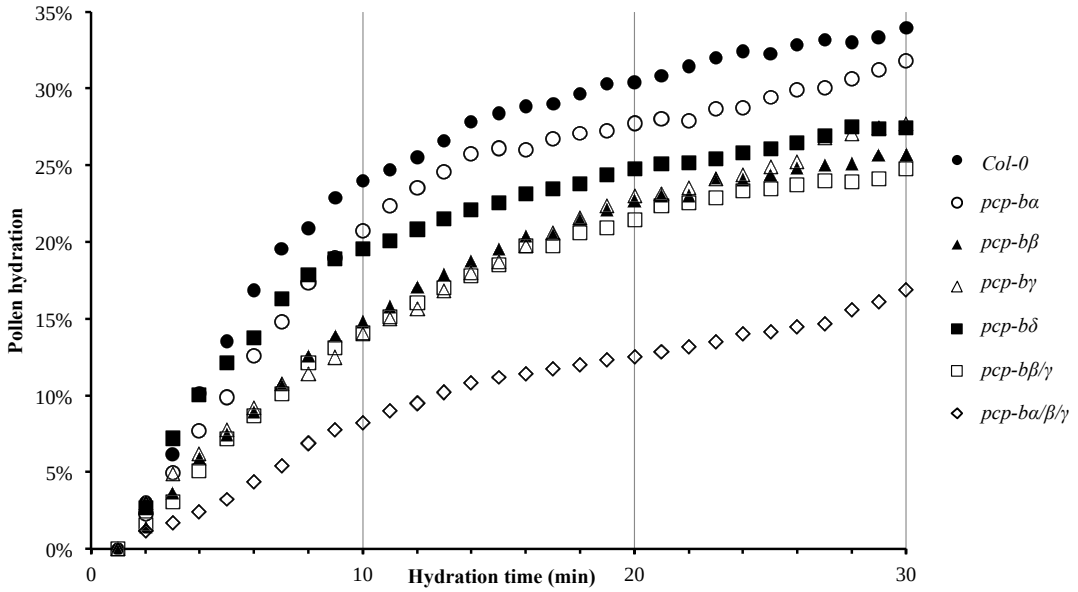
Fig. 3.5 Mutations in *Arabidopsis thaliana PCP-B* genes result in altered pollen hydration profiles. Box plots depict the 25% quartile, median, 75% quartile and full range of values. (a-d**)**, Pollen diameter distributions at 0 min, 10 min, 20 min and 30 min following pollination. 1.8 pixels = 1 μm. *, P<0.05, **, P<0.001, ***, P<0.0005 (Welsh's *t*-test). Sample sizes: Col-0, 16; *pcp-bα, pcp-bβ, pcp-bγ, pcp-bδ*, 15; *pcp-bα/β*, 16; *pcp-bα/β/γ*, 15. (e-g), Pollen hydration is represented as percentage change in pollen diameter relative to diameter at 0 mins (pollen diameter at initial contact with stigma) – distributions shown are for 10 min, 20 min and 30 min post-pollination. *, P<0.05, **, P<0.005, ***, P<0.000005 (Welsh's *t*-test). Sample sizes: Col-0, 16; *pcp-bα, pcp-bβ, pcp-bγ, pcp-bδ*, 15; *pcp-bα/β*, 16; *pcp-bα/β/γ*, 15.

We extended the analysis to determine the rate of pollen hydration on stigmas for mutant and wild-type pollen. Slopes were produced by linear regression based on pollen grain diameter during each 10-minute period following pollination. During the first ten-minute period wild-type pollen hydrates rapidly with this rate decreasing dramatically during the second and third ten-minute periods. A similar overall pattern was observed for all *pcp-b* mutant lines (Fig. 3.6a). During the first 10-minute period of pollination, the hydration rate of wild-type pollen was significantly higher than that for *pcp-bβ* and *pcp-bγ*. The *pcp-bβ/γ* double mutant also demonstrated a significantly affected hydration rate though this was not

greater than either single mutant. However introduction of *pcp-bα* into the *pcp-bβ/γ* line creating the *pcp-bα/β/γ* triple mutant had a dramatic effect on pollen hydration. *pcp-bα/β/γ* pollen hydrated at a substantially lower rate than either wild-type or pollen from the *pcp-bβ/γ* double mutant line (Fig. 3.6) despite the fact that the *pcp-bα* mutant had little discernible effect on pollen hydration in isolation. Hydration rates for the second and third 10-minute periods were not significantly different to wild-type for any of the mutant lines although the final extent of hydration was clearly lower for all lines with the exception of *pcp-bα*.

In order to determine if the pollen hydration defect resulted from an inherent inability of mutant pollen grains to absorb water rather than a defect in the pollen-stigma interaction, an *in vitro* pollen hydration assay was carried out. Using a humid chamber (providing 100% relative humidity) hydration of wild-type and *pcp-bα/β/γ* pollen was compared over a 30-minute period and their hydration characteristics were found to be indistinguishable (Fig. S1.7). Interestingly, comparison of wild-type pollen hydration on stigmas and in the humid chamber indicated that pollen hydrates more rapidly, and attains a greater degree of hydration on stigmas (Fig. 3.5b-d, S1.7b-d). These data demonstrate that the stigma is essential for rapid pollen hydration, and importantly, the absence of PCP-B protein from the pollen coat does not impair the biophysical ability of pollen to acquire water.

(a)



(b)

Pollen genotype	Sample size	Slope 1	Slope 2	Slope 3
Col-0	16	0.96±0.17	0.22±0.07	0.12±0.04
pcp-bα	15	0.84±0.18	0.19±0.07	0.16±0.06
pcp-bβ	15	**0.62±0.17***	0.27±0.09	0.11±0.04
pcp-bγ	15	**0.52±0.15*****	0.32±0.13	0.19±0.08
pcp-bδ	15	0.77±0.23	0.19±0.08	0.11±0.05
pcp-bβ/γ	16	**0.06±0.14****	0.26±0.10	0.10±0.03
pcp-bα/β/γ	15	**0.36±0.16******	0.14±0.05	0.17±0.03

Fig. 3.6 Rate of pollen hydration is severely decreased in *Arabidopsis thaliana pcp-b* triple mutants. (a) Curves of mean pollen hydration (% hydration is percentage change in pollen diameter). The vertical lines demark each ten-minute period over which slopes were calculated. (b) Rate of change in pollen diameter during the first three ten-minute periods of pollination. Average slopes ± confidence intervals were produced by linear regression. *, P<0.05, **, P<0.005, ***, P<0.001, ****, P<0.0005 (Welsh's *t*-test).

Pollen adhesion is reduced in *pcp-b* mutants

To further characterise the phenotype of *PCP-B* mutants, a pollen adhesion assay was devised which tested the ease with which pollen could be washed off the stigma. Significantly higher numbers of pollen grains from the *pcp-bα/β/γ* triple mutant (77%) were washed off wild-type stigmas compared to wild-type pollen (66%) 30 minutes post-pollination (Fig. 3.7a). However an EM ultrastructural analysis of the pollen from all *pcp-b* mutant lines revealed no discernible abnormalities in the characteristics of the pollen grain or pollen coat (Fig. 3.7b-g, S1.8, S1.9).

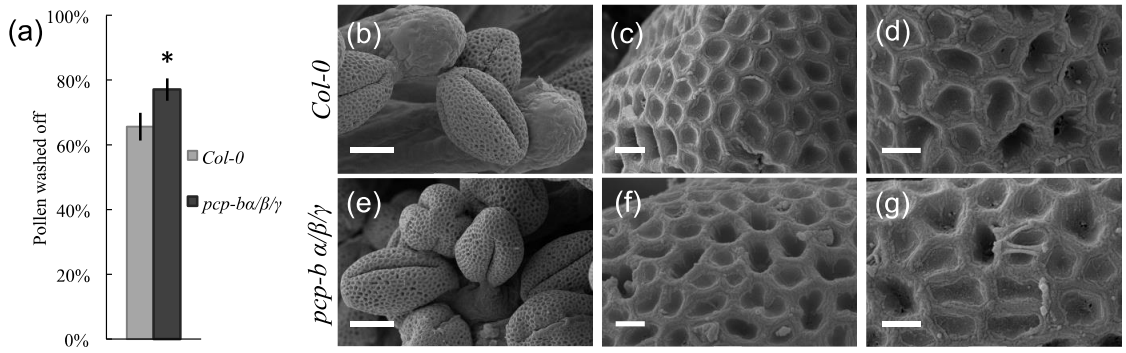


Fig. 3.7 Pollen morphology is unaffected in *Arabidopsis thaliana pcp-bα/β/γ* pollen grains and pollen-stigma adhesion is weakened. (a) The mean % of wild-type and *pcp-bα/β/γ* triple mutant pollen washed off wild-type stigmas in an adhesion assay 30 minutes post-pollination. Error bars represent the confidence interval. Sample sizes: wild-type stigmas, 64; triple mutant, 50. *P≪0.001 (Welsh's *t*-test). (b-g) Scanning Electron Microscopic (SEM) analysis of exine and pollen coat morphology in wild-type and *pcp-bα/β/γ* triple mutant plants. Scale bar: (b) and (e), 10µm. (c), (d), (f) and (g), 1µm. (b-d), wild-type. (e-g), *pcp-bα/β/γ*.

Initiation of pollen tube growth is delayed in *pcp-b* mutants

To determine if the early stages of pollen tube growth were affected by the delay in pollen hydration observed for *pcp-b* mutants, *in vivo* pollen tube lengths were estimated. After two hours wild-type pollen produced significantly longer tubes than pollen derived from all *pcp-b* mutant lines (Fig. 3.8a, S1.10) with this effect being largely maintained four hours post-pollination (Fig. 3.8b). This result is consistent with data collected from the pollen hydration assay where most mutants displayed impairment to the degree and rate of hydration which in turn would likely cause a delay in pollen tube emergence. Despite the observed post-pollination defects amongst the *pcp-b* mutants, there was no significant difference in seed set following self-pollinations compared with wild-type plants (Table S1.3), indicating that PCP-B protein function is likely restricted to very early post-pollination events.

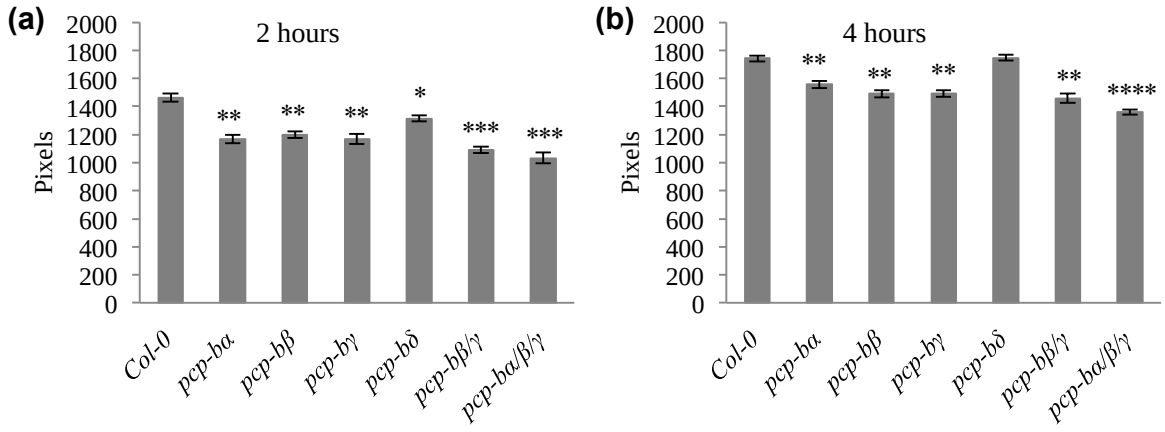


Fig. 3.8 Extent of *in vivo* pollen tube growth is reduced for *pcp-b* mutants. Distance (in pixels) of pollen tube growth for wild-type and *pcp-b* mutants two hours (a) and four hours (b) post-pollination. Pollen was applied to stigmas of the *A. thaliana* Col-0 A9-barnase male sterile line. Error bars represent the standard deviation. Sample sizes: 4. *, P<0.001, **, P<0.0001, ***, P<0.00001, ****, P<0.000001 (Welsh's *t*-test). 1 pixel = 0.625µm.

Structural prediction of AtPCP-Bs

Our analyses have revealed the presence of PCP-B-like proteins in a wide range of angiosperm lineages with all sequences sharing the characteristic motif of 8 cysteine residues in the mature polypeptide (Fig. 3.1). Costa *et al.* (2014) recently resolved the structure of the PCP-B-like protein ESF1.3 by nuclear magnetic resonance (NMR) and this made it possible to generate three-dimensional structural predictions for the *At*PCP-Bs by homologous alignment (Fig. 3.9, S1.11, S1.12). All resulting models were statistically well-supported (Table S1.4).

Based on the predicted three-dimensional structure *At*PCP-Bs likely share the same intramolecular disulphide bonding pattern as ESF1.3 (Fig. 3.9a) (Fig. 3.9b, S1.12) with all possessing a conserved cysteine-stabilised motif consisting of an α-helix and three-stranded antiparallel beta-sheet. In addition all *At*PCP-Bs have a conserved aromatic residue (Tyr-45 in *At*PCP-Bγ) that is also present in ESF1.3 (Trp-48) and other PCP-B-like proteins in *Arabidopsis thaliana* (Fig. 3.1). The surface electrostatic potential distribution for *At*PCP-Bγ (Fig. 3.9c) is characterised by both positively and negatively charged domains with a prominent positively charged extended loop held between Cys-36 and Cys-44 which lies in close proximity to the conserved aromatic residue (Tyr-45). These features are broadly shared between all four *At*PCP-Bs (Fig. 3.9, S1.12).

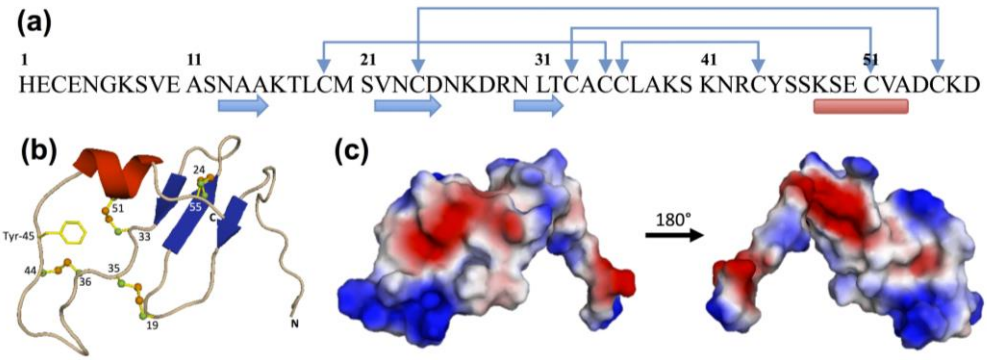


Fig. 3.9 AtPCP-Bγ structure prediction by SWISS-MODEL. (a) Amino acid sequence of PCP-Bγ. Connection arrows, disulphide bonds; Blue arrows, beta strands; Red bar, alpha helix. (b) Cartoon model of predicted structure of AtPCP-Bγ with indicated disulphide bonds and Tyrosine residue. (c) Distribution of electrostatic potential on AtPCP-Bγ surface based on the predicted structure. Blue, positive; Red, negative; white, hydrophobic residues.

Discussion

Compatible pollination is a highly regulated process that requires a suite of complementary pollen and pistil factors that act from the moment of pollen contact through to successful fusion of gametes (Edlund *et al.*, 2004; Hiscock & Allen, 2008). One of the earliest events in the establishment of compatibility amongst species that possess dry stigmas, such as *A. thaliana*, is an ability for pollen to gain access to stigmatic water (Elleman *et al.*, 1992; Safavian & Goring, 2013). This reproductive 'checkpoint' requires activation of a basal stigmatic compatibility system by factor(s) that must be derived from pollen (Safavian & Goring, 2013). Our investigations reported here into small pollen coating-borne cysteine-rich proteins point to an important role for the PCP-Bs in these earliest stages of pollen-pistil compatibility in Arabidopsis, as plants carrying mutations in *PCP-B* genes are impaired in pollen hydration. Importantly, PCP-Bs bear many hallmarks of intercellular signalling ligands and thus are likely to be a central component of a pollen molecular 'signature' that defines compatibility.

PCP-Bs are structurally related proteins that have an ancient evolutionary origin, being widespread amongst angiosperm taxa. Our phylogenetic analysis of 46 gene sequences encoding PCP-B-like proteins in *Arabidopsis* and *Brassica* (Fig. 3.3) reveals an evolutionary history featuring frequent gene duplication events and rapid sequence divergence around their conserved cysteine motif. These features are typical for gene

families associated with reproduction and importantly can contribute to reproductive isolation and speciation (Swanson & Vacquier, 2002; Clark *et al.*, 2006; Cui *et al.*, 2015). Interestingly the *PCP-Bs* investigated here were found to be closely related to the *ESF1*s that encode embryo developmental regulators (Costa *et al.*, 2014) and these sequences clustered in distinct phylogenetic clades, underlining their functional specialisation (Fig. 3.2, 3.3). Some *AtPCP-B* family members were more similar to genes in the closely related species *A. lyrata* suggesting that these have retained a specific function that predates speciation. For example *AtPCP-Bα* and *AtPCP-Bγ* are more closely related to the *Arabidopsis lyrata B4* and *B1* respectively than to other *AtPCP-Bs* (Fig. 3.3). Putative *Brassica* orthologues on the other hand were found in discrete clades more distant from the *Arabidopsis PCP-Bs* and could point to divergence of recognition factors required for compatibility. Species-specific functionalisation of plant reproductive proteins that contribute to reproductive isolation have been documented for the pollen tube attractant LURE proteins secreted by egg-accompanying synergid cells of the embryo sac (Takeuchi & Higashiyama, 2012). Heterologous expression of an *A. thaliana* LURE protein in *Torenia fournieri* synergid cells enabled *A. thaliana* pollen tubes to successfully locate and enter the embryo sac of this species. LURES are defensin-like CRPs and, in common with the PCP-B class proteins, are small secreted proteins that are encoded by a rapidly evolving gene family. It is thus tempting to speculate that PCP-Bs not only regulate aspects of compatibility but may also contribute to reproductive barriers within the Brassicaceae.

Our mutational study revealed that absence of *At*PCP-Bs from the pollen coat caused a series of interlinked phenotypes resulting from a primary defect in pollen hydration. We ascertained that the *pcp-b* hydration defect was not caused by gross morphological perturbation of the pollen coat (Fig. 3.7b-g, S1.8, S1.9) and that it was only evident during the pollen-stigma interaction, as triple mutant *pcp-bα/β/γ* pollen hydrated normally in a humid chamber (Fig. S1.7). Hydration rate, the degree of hydration and resulting pollen tube lengths were all found to be largely impaired amongst *pcp-b* single and combined mutants (Fig. 3.5, 3.6 and 3.8). We consider that the shorter tubes observed in pistils for *pcp-b* mutants is most likely the result of delayed pollen tube emergence rather than slower tube extension, as tube emergence is largely dependent on the degree of pollen hydration and pollen turgor (Taylor & Hepler, 1997). This inference was supported by the observation that triple mutant pollen adhered significantly less well to stigmatic papillae 30 minutes post-pollination (Fig. 3.7a) – we observed that a significant component of this

effect was due to wild-type pollen tubes initiating stigmatic penetration ahead of *pcp-b* pollen, thus anchoring them on the stigma, whereas substantially fewer mutant pollens had initiated germination (L. Wang and J. Doughty unpublished observation).

Comparison of the severity of the hydration defects between single and combined mutants revealed evidence of complex combinatorial effects of PCP-Bs in the pollen-stigma interaction. Out of the single mutant lines, *pcp-bγ* presented the most statistically robust hydration defect over the first 10-minute period following pollination, with *pcp-bβ* having an almost identical hydration profile (Fig. 3.5e, 3.6). Interestingly the phenotype of the double *pcp-bβ/γ* mutant was not additive, however when combined with the *pcp-bα* mutant, which singly had no significant phenotype, pollen hydration was dramatically reduced (Fig. 3.5e-g, 3.6). The contrasting combinatorial effects of these mutants suggests that PCP-Bα may be acting as a ligand to activate a different stigmatic hydration effector target to that of PCP-Bβ and PCP-Bγ, or that PCP-Bα acts to enhance activation of a putative stigmatic target working synergistically with other PCP-Bs. Similar complexity has been reported for synergid LURE proteins in *Torenia fournieri* and Arabidopsis where it seems likely that multiple LUREs work together, probably through different pollen tube receptors, to ensure appropriate pollen tube guidance to the embryo sac (Okuda *et al.*, 2009; Takeuchi & Higashiyama, 2012; Takeuchi & Higashiyama, 2016; Wang T *et al.*, 2016). The severity of the triple *pcp-bα/β/γ* mutant reduced the degree and rate of pollen hydration to almost one third that of wild-type and due to the close genetic linkage of *PCP-Bδ* to *PCP-Bγ* (<10 kb) we were unable to recover and test the effect of a *pcp-b* quadruple mutant. Thus it remains to be determined if a complete hydration block could be achieved by abolishing all PCP-B proteins from the pollen coat.

Given the structural features of *At*PCP-Bs and their homology to the ESF1 family of secreted developmental regulators we propose that PCP-Bs act as ligands to either directly or indirectly activate stigmatic targets that mediate transfer of water through the papilla plasma membrane. A substantial body of evidence now points to targeted stigmatic secretion as being a central feature of compatible pollination in both *A. thaliana* and *Brassica* and that the exocyst protein complex is essential to this process (Samuel *et al.*, 2009; Safavian & Goring, 2013; Safavian *et al.*, 2014; Safavian *et al.*, 2015). The exocyst mediates tethering of secretory vesicles to target membranes (Zarsky *et al.*, 2013) and stigmas from Arabidopsis that carry mutations in Exo70A1, a key linker component of the

exocyst tethering machinery, have severe pollen hydration defects. Targeted secretion likely delivers factors to the plasma membrane adjacent to compatible pollen that mediate water transport. For instance aquaporins, membrane-localised water transport proteins (Johanson *et al.*, 2001; Quigley *et al.*, 2002; Maurel *et al.*, 2008), could be deposited at the interface with compatible pollen. A specific role for pollen coat factors triggering such a response is supported by the observation that isolated *B. oleracea* pollen coat appears to evoke a secretory response by stigmatic papillae (Elleman & Dickinson, 1996).

Homology modelling of *At*PCP-Bs provided strong support for overall structural similarity with ESF1.3 (Fig. 3.9b, S1.12). As has been determined for ESF1.3 and other plant regulatory peptides it is likely that the disulphide-stabilised cysteine motif is crucial for protein function of PCP-Bs (Ohki *et al.*, 2011; Costa *et al.*, 2014). Intriguingly the *At*PCP-Bs shared a functionally essential aromatic residue with ESF1.3. Aromatic residues are a conserved feature of many plant regulatory peptides (Cao *et al.*, 2008; Okuda *et al.*, 2009; Sugano *et al.*, 2010; Costa *et al.*, 2012; Sprunck *et al.*, 2012) and are likely important in protein-protein interactions (Simpson *et al.*, 2000).

In conclusion this study shows that *At*PCP-Bs are important mediators of pollen hydration, a key early 'checkpoint' of pollen-stigma compatibility. Their close evolutionary relationship to the ESF1 family of embryo developmental regulators, and their broad similarity to other CRP regulatory proteins strongly suggest they act through interaction with as yet unknown stigmatic targets to activate the basal compatibility system. In addition, PCP-B maintenance and diversity within *Arabidopsis* and the Brassicaceae suggest that these proteins have the potential to contribute to prezygotic hybridization barriers.

Acknowledgments

We thank Susan Crennell for assistance with protein structural predictions, Ursula Potter for support with electron microscopy, Andrew James for plant maintenance, Jose Gutierrez-Marcos for the kind gift of *pAt5g61605:GUS* and *pAt2g16505:GUS* reporter lines, Tony Willis for protein sequencing, Sue Dixon for technical assistance, Hugh Dickinson for insightful discussions and Ed Nottingham, Hanis Jasmine Zaini and Ken Hayes for valuable contributions during their undergraduate projects. This research was

funded by a University of Bath postgraduate scholarship to LW and a Nuffield Foundation award to JD.

Author Contribution

JD and LW planned and designed the research and wrote the manuscript. LW was involved with all aspects of the research with LAC and RE contributing to expression analysis and CP to the SEM work. RJS and BQ assisted with experimental design and critical assessment of the manuscript.

References

Allen AM, Thorogood CJ, Hegarty MJ, Lexer C, Hiscock SJ. 2011. Pollen-pistil interactions and self-incompatibility in the Asteraceae: new insights from studies of *Senecio squalidus* (Oxford ragwort). *Ann Bot* **108**: 687-698.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen HM, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. 2003. Genome-wide Insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653-657.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195-201.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**: W252-W258.

Cao L, Bandelac G, Volgina A, Korostoff J, DiRienzo JM. 2008. Role of aromatic amino acids in receptor binding activity and subunit assembly of the cytolethal distending toxin of *Aggregatibacter actinomycetemcomitans*. *Infection and Immunity* **76**: 2812-2821.

Chae K, Kieslich CA, Morikis D, Kim SC, Lord EM. 2009. A gain-of-function mutation of Arabidopsis lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. *Plant Cell* **21**: 3902-3914.

Chapman LA, Goring, D.R. 2010. Pollen-pistil interactions regulating successful fertilization in the Brassicaceae. *Journal of Experimental Botany* **61**: 1987-1999.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* **131**: 11-22.

Costa LM, Marshall E, Tesfaye M, Silverstein KA, Mori M, Umetsu Y, Otterbach SL, Papareddy R, Dickinson HG, Boutiller K, et al. 2014. Central cell-derived peptides regulate early embryo patterning in flowering plants. *Science* **344**: 168-172.

Costa LM, Yuan J, Rouster J, Paul W, Dickinson H, Gutierrez-Marcos JF. 2012. Maternal control of nutrient allocation in plant Seeds by genomic imprinting. *Current Biology* **22**: 160-165.

Cui X, Lv Y, Chen ML, Nikoloski Z, Twell D, Zhang DB. 2015. Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Molecular Plant* **8**: 935-945.

Dickinson HG. 1995. Dry stigmas, water and self-incompatibility in Brassica. *Sexual Plant Reproduction* **8**: 1-10.

Dickinson HG, Elleman CJ, Doughty J. 2000. Pollen coatings - chimaeric genetics and new functions. *Sexual Plant Reproduction* **12**: 302-309.

Doughty J, Dixon S, Hiscock SJ, Willis AC, Parkin IAP, Dickinson HG. 1998. PCP-A1, a defensin-like Brassica pollen coat protein that binds the *S*-locus glycoprotein, is the product of gametophytic gene expression. *Plant Cell* **10**: 1333-1347.

Doughty J, Hedderson F, Mccubbin A, Dickinson H. 1993. Interaction between a coating-borne peptide of the Brassica pollen grain and stigmatic *S* (self-incompatibility)-locus-specific glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 467-471.

Doughty J, Wong HY, Dickinson HG. 2000. Cysteine-rich pollen coat proteins (PCPs) and their interactions with stigmatic *S* (incompatibility) and *S*-related proteins in Brassica: putative roles in SI and pollination. *Annals of Botany* **85**: 161-169.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Edlund AF, Swanson R, Preuss D. 2004. Pollen and stigma structure and function: the role of diversity in pollination. *Plant Cell* **16**: S84-S97.

Elleman CJ, Dickinson HG. 1986. Pollen-stigma interactions in Brassica. IV. Structural reorganization in the pollen grains during hydration. *Journal of Cell Science* **80**: 141-157.

Elleman CJ, Dickinson HG. 1990. The role of the exine coating in pollen-stigma interactions in *Brassica oleracea* L. *New Phytologist* **114**: 511-518.

Elleman CJ, Dickinson HG. 1996. Identification of pollen components regulating pollination-specific responses in the stigmatic papillae of *Brassica oleracea*. *New Phytologist* **133**: 197-205.

Elleman CJ, Franklin-Tong V, Dickinson HG. 1992. Pollination in species with dry stigmas: the nature of the early stigmatic response and the pathway taken by pollen tubes. *New Phytologist* **121**: 413-424.

Fiebig A, Mayfield JA, Miley NL, Chau S, Fischer RL, Preuss D. 2000. Alterations in *CER6*, a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* **12**: 2001-2008.

Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30**: S162-S173.

Hernandez-Pinzon I, Ross JH, Barnes KA, Damant AP, Murphy DJ. 1999. Composition and role of tapetal lipid bodies in the biogenesis of the pollen coat of *Brassica napus*. *Planta* **208**: 588-598.

Heslop-Harrison Y, Shivanna KR. 1977. The receptive surface of the angiosperm stigma. *Annals of Botany* **41**: 1233-1258.

Hiscock SJ, Allen AM. 2008. Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus. *New Phytologist* **179**: 286-317.

Hiscock SJ, Doughty J, Willis AC, Dickinson HG. 1995. A 7-kDa pollen coating-borne peptide from *Brassica napus* interacts with *S*-locus glycoprotein and *S*-locus-related glycoprotein. *Planta* **196**: 367-374.

Hiroi K, Song M, Sakazono S, Osaka M, Masuko-Suzuki H, Matsuda T, Suzuki G, Suwabe K, Watanabe M. 2013. Time-lapse imaging of self- and cross-pollination in *Brassica rapa*. *Annals of Botany* **112**: 115-122.

Hulskamp M, Kopczak SD, Horejsi TF, Kihl BK, Pruitt RE. 1995. Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. *Plant Journal* **8**: 703-714.

Johanson U, Karlsson M, Johansson I, Gustavsson S, Sjovall S, Fraysse L, Weig AR, Kjellbom P. 2001. The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiology* **126**: 1358-1369.

Kho YO, Baer J. 1968. Observing pollen tubes by means of fluorescence. *Euphytica* **17**: 298-302.

Kiefer F, Arnold K, Kuenzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research* **37**: D387-D392.

Kleinboelting N, Huep G, Kloetgen A, Viehoever P, Weisshaar B. 2012. GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Research* **40**: D1211-D1215.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127-128.

Li YL, Dai XR, Yue X, Gao XQ, Zhang XS. 2014. Identification of small secreted peptides (SSPs) in maize and expression analysis of partial *SSP* genes in reproductive tissues. *Planta* **240**: 713-728.

Ma J-F, Liu Z-H, Chu C-P, Hu Z-Y, Wang X-L, Zhang XS. 2012. Different regulatory processes control pollen hydration and germination in *Arabidopsis*. *Sexual Plant Reproduction* **25**:77-82.

Marshall E, Costa LM, Gutierrez-Marcos J. 2011. Cysteine-Rich Peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *Journal of Experimental Botany* **62**: 1677-1686.

Maurel C, Verdoucq L, Luu DT, Santoni V. 2008. Plant aquaporins: membrane channels with multiple integrated functions. *Annual Review of Plant Biology* **59**: 595-624.

Mayfield JA, Preuss D. 2000. Rapid initiation of Arabidopsis pollination requires the oleosin-domain protein GRP17. *Natural Cell Biology* **2**: 128-130.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**: 205-217.

Ohki S, Takeuchi M, Mori M. 2011. The NMR structure of stomagen reveals the basis of stomatal density regulation by plant peptide hormones. *Nature Communications* **2**: 512.

Okuda S, Tsutsui H, Shiina K, Sprunck S, Takeuchi H, Yui R, Kasahara RD, Hamamura Y, Mizukami A, Susaki D, et al. 2009. Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**: 357-361.

Paul W, Hodge R, Smartt S, Draper J, Scott R. 1992. The isolation and characterisation of the tapetum-specific *Arabidopsis thaliana* A9 gene. *Plant Molecular Biology* **19**: 611-622.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785-786.

Piffanelli P, Ross JHE, Murphy DJ. 1998. Biogenesis and function of the lipidic structures of pollen grains. *Sexual Plant Reproduction* **11**: 65-80.

Preuss D, Lemieux B, Yen G, Davis RW. 1993. A conditional sterile mutation eliminates surface components from Arabidopsis pollen and disrupts cell signaling during fertilization. *Genes & Development* **7**: 974-985.

Quigley F, Rosenberg JM, Shachar-Hill Y, Bohnert HJ. 2002. From genome to function: the Arabidopsis aquaporins. *Genome Biol* **3**: RESEARCH0001.

Safavian D, Goring DR. 2013. Secretory activity is rapidly induced in stigmatic papillae by compatible pollen, but inhibited for self-incompatible pollen in the Brassicaceae. *Plos ONE* **8**: e84286.

Safavian D, Jamshed M, Sankaranarayanan S, Indriolo E, Samuel MA, Goring DR. 2014. High humidity partially rescues the *Arabidopsis thaliana exo70A1* stigmatic defect for accepting compatible pollen. *Plant Reproduction* **27**: 121-127.

Safavian D, Zayed Y, Indriolo E, Chapman L, Ahmed A, Goring, DR. 2015. RNA silencing of exocyst genes in the stigma impairs the acceptance of compatible pollen in *Arabidopsis*. *Plant Physiology* **169**: 2526-2538.

Samuel MA, Chong YT, Haasen KE, Aldea-Brydges MG, Stone SL, Goring DR. 2009. Cellular pathways regulating responses to compatible and self-incompatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. *Plant Cell* **21**: 2655-2671.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671-675.

Schopfer CR, Nasrallah ME, Nasrallah JB. 1999. The male determinant of self-incompatibility in *Brassica*. *Science* **286**: 1697-1700.

Shiba H, Takayama S, Iwano M, Shimosato H, Funato M, Nakagawa T, Che FS, Suzuki G, Watanabe M, Hinata K, et al. 2001. A pollen coat protein, SP11/SCR, determines the pollen *S*-specificity in the self-incompatibility of *Brassica* species. *Plant Physiology* **125**: 2095-2103.

Silverstein KAT, Moskal WA, Jr., Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA. 2007. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant Journal* **51**: 262-280.

Simpson PJ, Xie HF, Bolam DN, Gilbert HJ, Williamson MP. 2000. The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *Journal of Biological Chemistry* **275**: 41137-41142.

Smith AG, Eberle CA, Moss NG, Anderson NO, Clasen BM, Hegeman AD. 2013. The transmitting tissue of *Nicotiana tabacum* is not essential to pollen tube growth, and its ablation can reverse prezygotic interspecific barriers. *Plant Reproduction* **26**: 339-350.

Smyth DR, Bowman JL, Meyerowitz EM. 1990. Early flower development in *Arabidopsis*. *Plant Cell* **2**(8): 755-767.

Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U, Dresselhaus T. 2012. Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* **338**: 1093-1097.

Stone SL, Anderson EM, Mullen RT, Goring DR. 2003. ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. *Plant Cell* **15**: 885-898.

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, Mori M, Hara-Nishimura I. 2010. Stomagen positively regulates stomatal density in *Arabidopsis*. *Nature* **463**: 241-244.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**: 137-144.

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. 2000. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. *Nature* **403**: 913-916.

Takayama S, Shiba H, Iwano M, Asano K, Hara M, Che FS, Watanabe M, Hinata K, Isogai A. 2000. Isolation and characterization of pollen coat proteins of *Brassica campestris* that interact with *S* locus-related glycoprotein 1 involved in pollen-stigma adhesion. *Proceedings of the National Academy of Sciences, USA* **97**: 3765-3770.

Takeuchi H, Higashiyama T. 2012. A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *Plos Biology* **10**: e1001449.

Takeuchi H, Higashiyama T. 2016. Tip-localized receptors control pollen tube growth and LURE sensing in *Arabidopsis*. *Nature* **531**: 245-248.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Taylor LP, Hepler PK. 1997. Pollen germination and tube growth. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**: 461-491.

Updegraff EP, Zhao F, Preuss D. 2009. The extracellular lipase EXL4 is required for efficient hydration of *Arabidopsis* pollen. *Sexual Plant Reproduction* **22**: 197-204.

Villar M, Gaget M, Said C, Knox RB, Dumas C. 1987. Incompatibility in *Populus*: structural and cytochemical characteristics of the receptive stigmas of *Populus alba* and *Populus nigra*. *Journal of Cell Science* **87**: 483-490.

Wang T, Liang L, Xue Y, Jia PF, Chen W, Zhang MX, Wang YC, Li HJ, Yang WC. 2016. A receptor heteromer mediates the male perception of female attractants in plants. *Nature* **531**(7593): 241-244.

Yost JM, Kay KM. 2009. The evolution of postpollination reproductive isolation in *Costus*. *Sexual Plant Reproduction* **22**(4): 247-255.

Zarsky V, Kulich I, Fendrych M, Pecenkova T. 2013. Exocyst complexes multiple functions in plant cells secretory pathways. *Current Opinion in Plant Biology* **16**: 726-733.

(End of manuscript)

3.3 Discussion

For decades, studies have been focusing on the molecular basis of self-incompatibility (SI) in angiosperms, whereas the establishment of pollen-stigma compatibility is still poorly understood. Pollen hydration can be seen as the first compatibility 'checkpoint' in plant reproduction in species possessing 'dry' stigmas. Although it is now well established that pollen compatibility involves exocyst-mediated stigmatic secretion at the contact point with pollen (at least in the Brassicaceae), which likely permits water transport out of papillar cells (Samuel *et al.*, 2009; Safavian & Goring, 2013), the earliest step of pollen-stigma recognition for this event remained unknown. The work reported in this chapter demonstrated that PCP-B pollen coat proteins are key regulators of pollen hydration, and this represents the very first example of a pollen surface protein acting as mobile ligand that regulates compatibility by modulating hydration.

Interestingly, three PCP-B-like proteins, ESF1.1, 1.2 and 1.3, were shown as being highly homologous to AtPCP-B. They have recently been characterised as playing a signalling role during the early stages of embryogenesis in Arabidopsis (Costa *et al.*, 2014). By utilising mutants with individually disrupted disulphide bonds, Costa *et al.* (2014) revealed that ESF1.3 structural topology is essential for its biological activity and that this is supported by four disulphide bonds. Full resolution of the ESF1.3 (At1g10717) three-dimensional structure made its sequence an ideal homologous alignment template for predicting AtPCP-B structures and potential functional sites. Interestingly, similar to members of the plant defensin-like protein class of CRPs, AtPCP-B predicted structures include a cysteine-stabilised α/β motif, which has been observed in multiple plant antimicrobial proteins (AMPs) and has been considered to be important for antimicrobial activity (Bruix *et al.*, 1993; Fant *et al.*, 1998; Fant *et al.*, 1999; Sagaram *et al.*, 2013). It has also been described that the positively charged residues in loops and β-sheet regions are crucial for antifungal activity of plant defensins (Fant *et al.*, 1998). The structure-function evidence of plant AMPs and the identical and unique disulphide bond-forming pattern in ESF1.3 and AtPCP-Bs suggests that AtPCP-Bs may act as signalling molecules by electrostatically interacting with targets on stigma papillar cells. In the following chapter, more work was carried out to investigate potential interaction targets of AtPCP-Bs. Further structure-function studies will also need to be carried out in the future to identify precisely

which regions of PCP-Bs are crucial for activation of the as-yet-uncharacterised downstream ‘signalling’ pathway that functions in the earliest stages of self-compatibility.

Chapter 4 Heterologous expression and putative stigmatic binding targets of AtPCP-Bs

4.1 Introduction

Sexual reproduction in angiosperms requires molecular recognition between male and female reproductive tissues. The entire process starts from the initial point of pollination and lasts through events of pollen-pistil interactions where compatibility is established through factors that mediate guidance of the pollen tube to the ovule (reviewed in Hiscock & Allen, 2008). A class of pollen coat cysteine-rich proteins (CRPs), the AtPCP-Bs, have been identified as important regulators of the pollen-stigma interaction acting to regulate pollen hydration (Chapter 3, Wang L *et al*., 2016). Previously identified pollen coat CRPs have been found to interact with stigmatic targets (Chapter 1, 1.3.3.1). For example, in self-incompatible *Brassica* species, the *S*-locus cysteine-rich protein (SCR/SP11) acts as the male determinant, interacting with the stigmatic female determinant, *S*-receptor kinase (SRK) (Schopfer *et al.*, 1999; Takasaki *et al.*, 2000; Shiba *et al.*, 2001). A member of the PCP-A class of pollen coat proteins, PCP-A1, was identified by its ability to bind the stigmatic S-locus glycoprotein (SLG) in *Brassica oleracea* (Doughty *et al.*, 1998). Despite strong circumstantial evidence suggesting that these PCP-B class small cysteine-rich proteins (CRPs) are acting as signalling molecules in the basal compatibility system (Chapter 3, Wang L *et al*., 2016), the identity of the putative stigmatic targets and the mechanism by which PCP-Bs influence pollen hydration remains unknown.

The search for novel protein-protein interactions in plants is challenging. Although extensive work has been done on plant signalling molecules during the last several decades, very few ligand-receptor pairs have been identified. These signalling molecules have been found to be involved in a variety of processes. The interactions of Systemins with SR160 (systemin receptor) / BRI1 (Brassinolide receptor kinase 1), were found in the induction of the plant wound response (Meindl *et al*., 1998; Scheer & Ryan 2002). The binding between the CLAVATA 3 (CLV3) and CLV1/CLV2 heterodimers has been discovered to regulate shoot meristem development in Arabidopsis (Trotochaud *et al*., 2000). As described in Chapter 1, multiple CRPs act as signalling ligands interacting with their receptor to mediate intercellular communication, including SCR-SRK and PrsS-PrpS regulating self-

incompatibility system (see section 1.3.3.1), RALF-FERONIA regulating root development, EFP-TMM in specifying stomatal patterning (see section 1.3.1) and LURE-MDIS1/MIK regulating pollen tube guidance (see section 1.3.3.2), etc. The common methods are largely limited to affinity purification (e.g. crosslinking and pull-down assays) and heterologous *in vivo* systems (e.g. yeast-two-hybrid). However, these approaches are frequently prone to generating false positives, especially for proteins that interact transiently (reviewed in Qi & Katagiri, 2009), thus the discovery of novel binding must be validated by multiple methods. The very recent discovery of interactions between the pollen tube attractant LURE and its pollen tube heteromer receptor MDIS1-MIK demonstrated the difficulties in identifying interactors for 'orphan' ligands or receptors (Wang T *et al.*, 2016). The search for the male receptor was hampered not only by the complexity of the interaction mechanism but also by the redundancy of LUREs, a situation which may also be applied to this study.

In order to gain an insight into the identity of putative PCP-B receptors, *At*PCP-Bs were heterologously expressed with a view to utilising purified *At*PCP-Bs as 'bait' in a series of protein-protein interaction assays. Using membrane proteins and total cell proteins from Arabidopsis stigmas, a series of protein-protein interaction assays were carried out. These included far-western blotting as well as two affinity purification assays (protein crosslinking and pull-down) combined with mass spectrometry (AP-MS).

This work identified a number of stigmatic PCP-B binding proteins as putative *At*PCP-B 'receptors'. Future work will need to be carried out to validate these putative *At*PCP-B targets and explain the signalling mechanism by which they could act to regulate pollen hydration.

4.2 Results

4.2.1 Cloning of *AtPCP-B* genes

Heterologously expressed AtPCP-Bs were planned to be utilised in a series of protein-protein interaction studies. The expression of cysteine-rich proteins is a challenge due to the prerequisite that all intramolecular disulphide bonds are formed in the correct pattern and the proteins have good solubility. To obtain native-state AtPCP-Bs in large quantities, multiple factors have been considered such as the inclusion of fusion tags, choice of a suitable *E.coli* strain and the specific expression conditions. The choice of protein expression vector pET-32a was based on the gene cloning strategy, ease of protein purification and downstream applications. The expression vector pET-32a includes His-tag and S-tag purification / immobilization motifs and a thioredoxin (Trx)-tag (Appendix 2.3). The His-tag is essential for purification and immobilisation of the recombinant proteins and can be readily detected by commercially available antibodies. The S-tag improves protein solubility and the Trx-tag can facilitate proper folding by facilitating the reduction of proteins (Yasukawa *et al.*, 1995). The fusion tags can be removed by site-specific protease digestion with thrombin and enterokinase. The choice of *E.coli* strain, Origami^{TM} 2, was important to facilitate protein folding. This cell line carries mutations in thioredoxin and glutathione reductase which provides an oxidative environment in the cell, thus facilitating the formation of disulphide bonds between cysteine residues of overexpressed proteins (Stewart *et al.*, 1998).

To express the AtPCP-Bs in *E.coli*, the AtPCP-B encoding genes first had to be polymerase chain reaction (PCR) amplified and subsequently cloned into an expression vector. Based on the expression analysis for PCP-Bs (Chapter 3, Wang L *et al*., 2016), transcripts were found to be maximal in stage 12 anthers. Thus, mRNA was isolated from stage 12 anthers for reverse-transcriptase PCR (RT-PCR). cDNA synthesis products from stage 12 anthers were verified by PCR with *GapC* (*At3g04120*) primers (Appendix 2 Figure S4.1). *GapC* is a ubiquitously expressed gene in *Arabidopsis thaliana*, encoding the glyceraldehyde-3-phosphate dehydrogenase C-subunit. A genomic DNA sample was used as template in the positive control whereas water was added to the negative control for excluding false positive results caused by contamination. The PCR products indicated the presence of *GapC* in the cDNA products. The *GapC* primers span the intron of genomic DNA, thus the size of the PCR product from Arabidopsis leaf gDNA is larger than the

products from the anther-derived cDNA. This result confirmed that cDNA synthesis from stage 12 anthers was successful and that there was no significant gDNA contamination in these samples that would be used for subsequent cloning of the *PCP-B*s (Appendix 2 Figure S4.1).

The primer sequences designed for amplifying *AtPCP-B* genes can be found in Appendix 2 (Table S4.1). Restriction enzyme sites were included in the primers to permit directional cloning of the *PCP-B* gene sequences encoding mature proteins into the cloning vectors (pGEM®-T easy and pJET1.2) and the expression vector (pET-32a). The Tm of each primer and the predicted PCR product size for each primer pair can be found in Table 4.1. The RT-PCR products generated were of the predicted sizes indicating that the four *AtPCP-B* genes were amplified successfully (Figure 4.1).

Table 4.1 | Primers used for *AtPCP-B* gene amplification

Gene code	Primers	Tm/°C	Predicted size
At5g61605	AtPCP-Bα F	70	165
At5g61605	AtPCP-Bα R	72	
At2g29790	AtPCP-Bβ F	67	156
At2g29790	AtPCP-Bβ R	68	
At2g16535	AtPCP-Bγ F	66	162
At2g16535	AtPCP-Bγ R	66	
At2g16505	AtPCP-Bδ F	64	165
At2g16505	AtPCP-Bδ R	68	

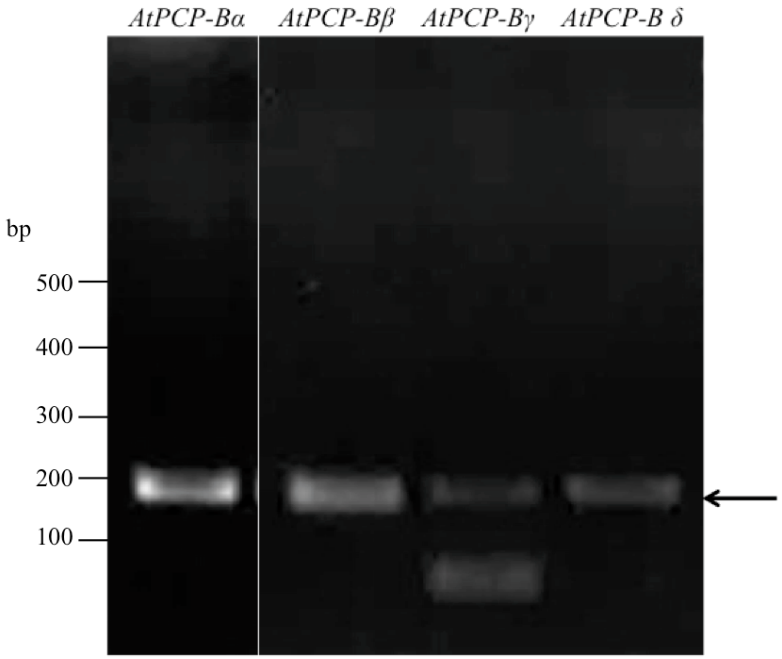


Figure 4.1 | RT-PCR products of *AtPCP-B* genes amplified from stage 12 anther cDNA. 1% agarose gel stained with ethidium bromide. The arrow indicates amplified products that are consistent with the predicted sizes for the *AtPCP-B* target sequences encoding mature PCP-B proteins.

The target genes were cloned into a PCR cloning vector first and then moved to the expression vector. For the cloning vector, *AtPCP-Bβ, γ,* and *δ* were cloned into pGEM®-T easy vector whilst *AtPCP-Bα* was cloned using the pJET1.2 vector. All four target genes were successfully cloned into the pET-32a expression vector (Appendix 2.3) for the production of recombinant fusion proteins. Cloning products were initially validated by colony PCR confirming the presence of the target DNA in the recombinant vectors. Subsequent sequencing and BLAST® alignment analysis (Appendix 2.4) of the cloned products confirmed the presence of target gene insertion and that the ORFs were maintained for correct fusion protein expression.

4.2.2 Confirming alternative splicing of *At2g29790/AtPCP-Bβ*

In *Arabidopsis thaliana*, approximately 42% of intron-containing genes are alternative spliced (Filichkin *et al.*, 2010). Preliminary results obtained from previous research in the lab and a recently published study (Costa *et al.*, 2014) indicated two potential gene models for *At2g29790* (Figure 4.2), with both encoding a PCP-B-like protein. However, there was

no definitive empirical evidence supporting the gene model prediction that *At2g29790* makes a PCP-B-like transcript. To confirm whether *At2g29790* was alternatively spliced as we hypothesised, four pairs of primers were designed and utilised for scanning corresponding transcripts from cDNA (Table 4.2 and Figure 4.2). Based on the preliminary RT-PCR with these primers, primer F1, which corresponds to a region of *AtPCP-Bβ* exon 2, failed to amplify a product. Except for the RT-PCR product generated by primers F2 and R1, (which were previously used for amplifying *AtPCP-Bβ* for heterologous expression thus confirming this cDNA sequence), the other five RT-PCR products (Figure 4.3) were cloned into the pJET1.2 vector for DNA sequencing. DNA alignment of the resulting sequencing data (Appendix 2.4) provided a consensus for the *AtPCP-Bβ* transcript (Figure 4.3). Although RT-PCR using forward primer 'F1' (that corresponds to *AtPCP-Bβ* exon 2) did not successfully amplify the spanning cDNA fragment, the *AtPCP-Bβ* exon 2 sequence was present in the cloned RT-PCR product generated by primers 'F3' and 'F4', which correspond to a region upstream of *AtPCP-Bβ* exon 2 (Figure 4.2, 4.3). The pooled sequence data from this series of experiments indicates that exon 2 of *AtPCP-Bβ* is likely to be transcribed. In sum, the integrated sequence strictly corresponds to *AtPCP-Bβ* exons 1, 2, 3, and 4 (Figure 4.2). The assembled sequence was translated into all three ORFs (Appendix 2.5) and reading frame 'ORF1' was found to encode a continuous polypeptide with the characteristic eight-cysteine residue motif region of the PCP-B class CRPs. Thus, the 'ORF1' is likely to be the functional reading frame of the *AtPCP-Bβ*. These data confirm that *AtPCP-Bβ* encoding transcripts differ from those reported for *At2g29790* and thus confirm that alternative slicing occurs in *At2g29790.*

Table 4.2 | Primers used for the study of *AtPCP-Bβ* (*At2g29790*) alternative splicing

Name	Primer direction	Primer sequences	Tm/°C
F1	Forward	5'-GACAAGACACACAGTTTTGACG	55
F2	Forward	5'-CAAGGTTTTGACGCGAACAAG	55
F3	Forward	5'-GTCATCATCACATTTTGTTATCCTTTG	53
F4	Forward	5'-TTCGGATTTCAACCATCAAACATAAAC	54
R1	Reverse	5'-ATGAACAAGCCGGTTGCAG	57
R2	Reverse	5'-TTGTAGCTGTATTGTTTGTATAAGGTC	53

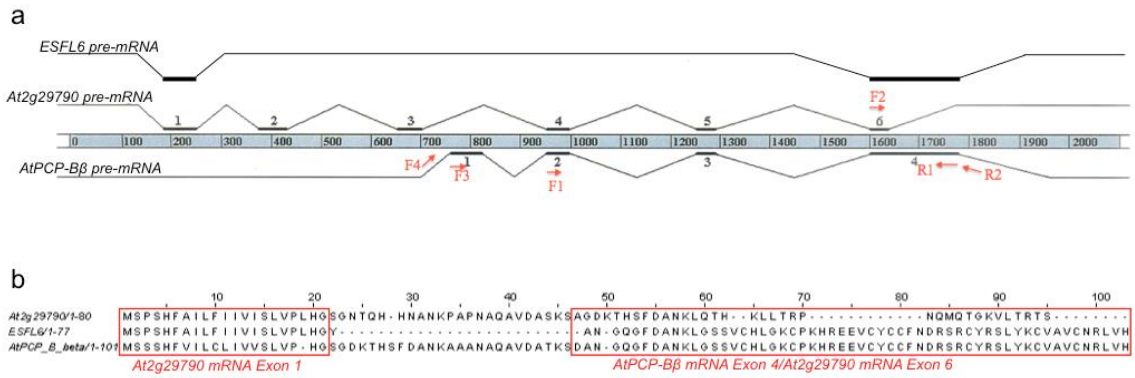
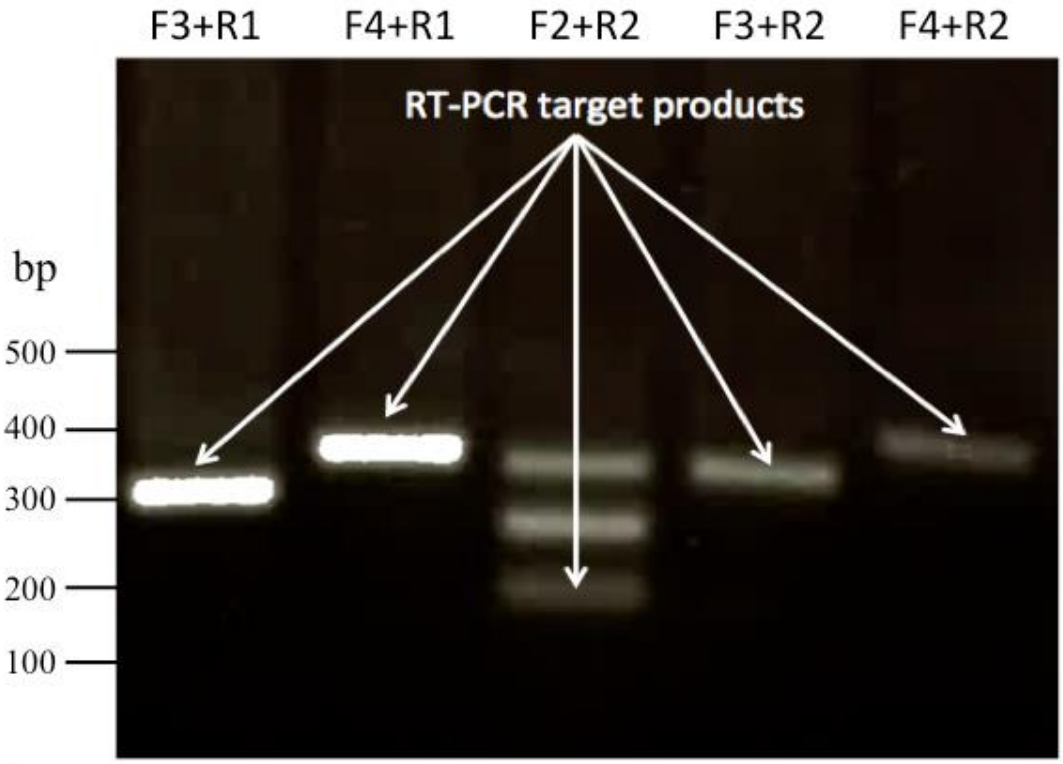


Figure 4.2 | Schematic diagram of *At2g29790* alternative splicing (a) and the alignment of protein sequences translated from each mRNA (b). The horizontal black bars indicate the exons processed into mature mRNAs (detailed DNA sequence see Appendix 2.7). The red arrows show the primers designed for alternative splicing confirmation PCR. The red boxes highlight the exons encoding EMBRYO SURROUNDING FACTOR 1-like protein 6 (ESFL6).



	R1			R2		
	Sample	Predict size	Appear size	Sample	Predicted size	Actual size
F1	No product					
F2	Gene cloning for heterologous expression			4	192	~190
F3	2	297	~300	5	337	~340
F4	3	354	~350	6	390	~390

Figure 4.3 | RT-PCR products for *AtPCP-Bβ* providing evidence for alternative splicing at the *At2g29790* locus. The primers used in each sample are shown above the gel lanes.

4.2.3 Expression and purification of recombinant AtPCP-Bs from *E.coli*

To discover the potential stigmatic targets of AtPCP-Bs, the four AtPCP-Bs were cloned and expressed in *E.coli* as fusion proteins for protein-protein interaction assays. A schematic of the predicted fusion proteins is shown in Figure 4.4 (and Appendix 2.6). Protein extracts were prepared from the soluble fraction, insoluble fraction and as well as a total cell protein sample (TCP) from each 100 ml trial expression cultures. 10% of the extracted protein from 1ml of cell culture was analysed by SDS-PAGE. Based on the

proportion of soluble PCP-B fusion protein produced by various culture conditions (Figure 4.5, 4.6), expression was optimised for each of the four constructs such that the best yield of soluble protein was obtained for subsequent purification (Table 4.3). Induction of expression, culture temperature and time of culture were all varied in the optimisation screen. Expression of all four target proteins was induced effectively with 0.4mM IPTG. The expression of AtPCP-Bα, γ, and δ was carried out for four hours at 37°C, while the expression of AtPCP-Bβ was carried out for twenty hours at 16°C. Under the optimised conditions, the proportion of the soluble fraction for AtPCP-Bα, γ, and δ achieved at least 50%, whilst the proportion of soluble fraction for AtPCP-Bβ achieved at least 10%.

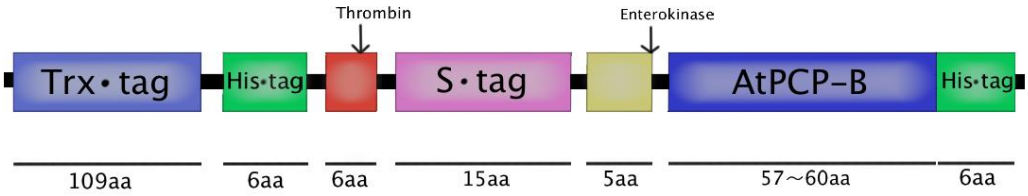


Figure 4.4 | Schematic diagram of PCP-B fusion proteins produced by the pET32a expression vector. Arrows indicate the available cleavage sites for thrombin and enterokinase respectively. Length of tags and enzyme recognition sites are labelled below the diagram. Black bars indicate sequences between the major regions of significance.

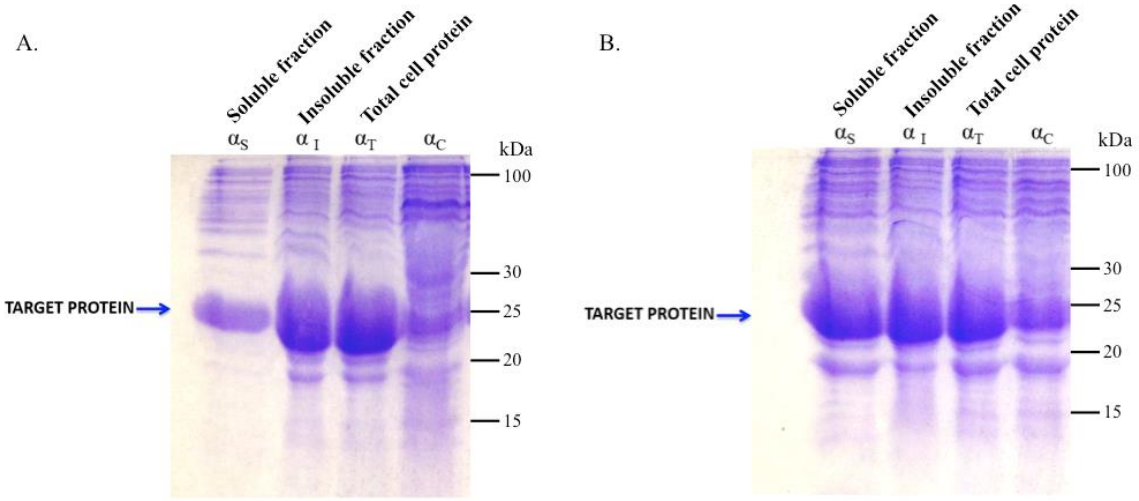


Figure 4.5 | Analysis of recombinant AtPCP-Bα protein expression by SDS-PAGE. The SDS-PAGE gels (18%) with soluble fraction (α_S), insoluble fraction (α_I) and total cell protein (α_T) of expressed pET-32a-recombinant AtPCP-Bα in *E.coli* (Origami™ 2 (DE3)) after 0.4mM IPTG induction, as well as the control (α_C) without induction treatment. Expression conditions: 20 hours at 16°C (A) and 4 hours at 37°C (B).

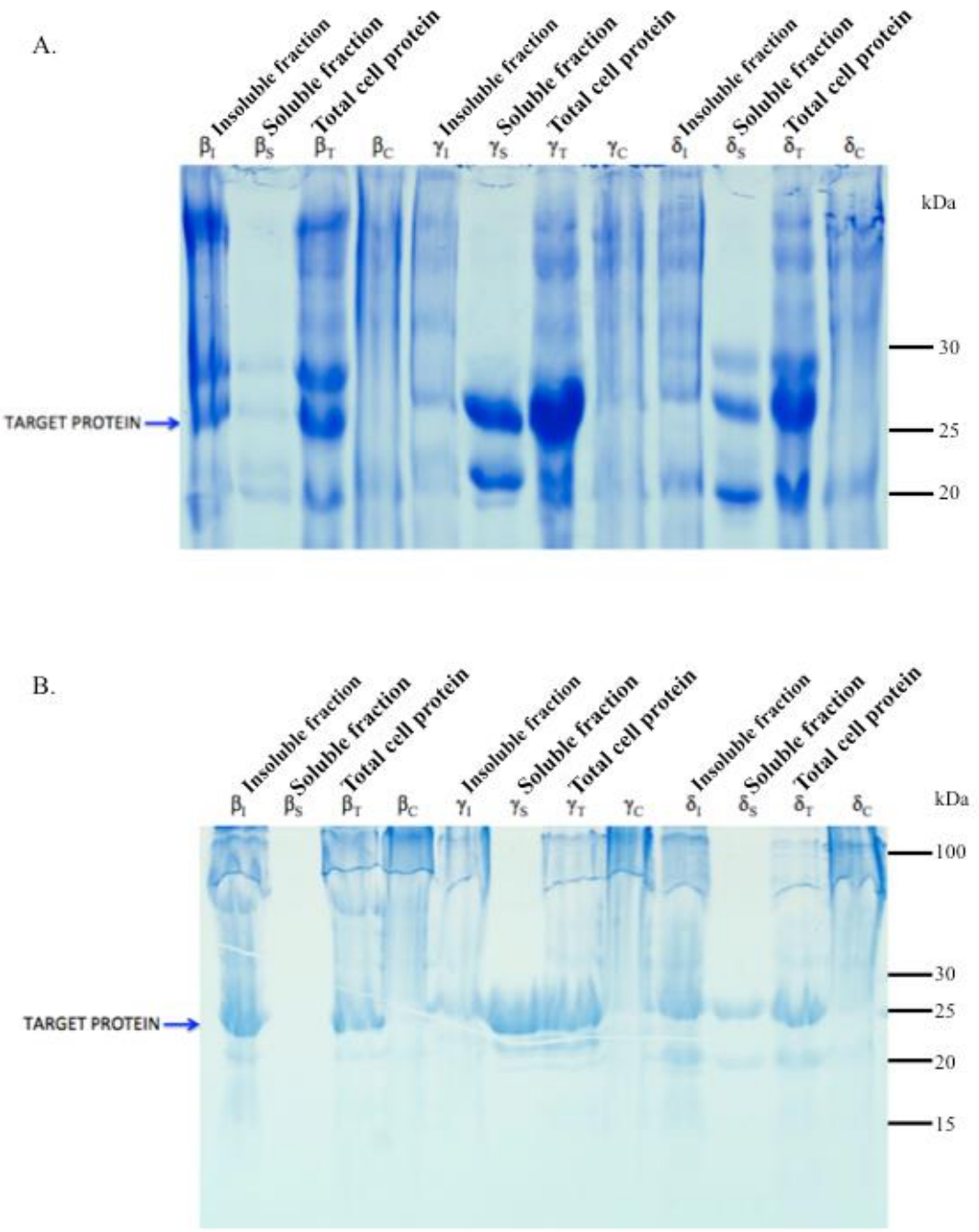


Figure 4.6 | Analysis of recombinant AtPCP-B protein expression by SDS-PAGE. SDS-PAGE gel (18%) with total cell protein (β_T, γ_T, δ_T), soluble fraction (β_S, γ_S, δ_S) and insoluble fraction (β_I, γ_I, δ_I) of expressed pET-32a-recombinant AtPCP-Bβ, γ in *E.coli* (Origami™ 2 (DE3)) after induction, as well as the control (β_C, γ_C, δ_C) without induction treatment. Expression conditions: 20 hours at 16°C (A) and 4 hours at 37°C (B).

Table 4.3 | Optimised expression conditions for each *At*PCP-B target protein

Protein	IPTG concentration	Culture temperature	IPTG induction time
AtPCP-Bα	0.4mM	37°C	4 hours
AtPCP-Bβ	0.4mM	16°C	20 hours
AtPCP-Bγ	0.4mM	37°C	4 hours
AtPCP-Bδ	0.4mM	37°C	4 hours

To produce soluble AtPCP-Bs for protein-protein interaction assays, the fusion proteins were first purified from the soluble fraction following cell lysis (Figure 4.7) by utilising immobilised metal-affinity chromatography (IMAC) based on the interaction between Ni^{2+} and the histidine imidazole ring. The N-terminal tags were then cleaved by enterokinase to yield the mature PCP-B protein linked to a His-tag at C-terminal (His-PCP-B) (Fig 4.8). Enterokinase treatment yielded a protein solution containing cleaved tags, uncleaved recombinant protein and cleaved His-PCP-B that could be utilised in protein-protein interaction assays. Due to the presence of the His-tag on both the mature PCP-B (His-PCP-B) and the cleaved N-terminal region of the recombinant protein, the mature AtPCP-Bs were isolated by reverse phase chromatography (RPC). The purified recombinant AtPCP-Bs and mature AtPCP-Bs were verified by western-blotting (Figure 4.9) which confirmed the presence of the target proteins. It is worth noting that despite a single clean polypeptide band being present for AtPCP-Bβ following initial purification of the fusion protein a double band was observed on gels following enterokinase cleavage. This is most likely due to partial refolding/disulphide bond formation for AtPCP-Bβ resulting in differing mobilities of the protein on the gel. Alternatively it cannot be ruled out that there is some non-specific cleavage of the protein by enterokinase. Nevertheless, heterologously expressed AtPCP-Bs with relatively high purity and native-state were obtained for further protein-protein interaction assays.

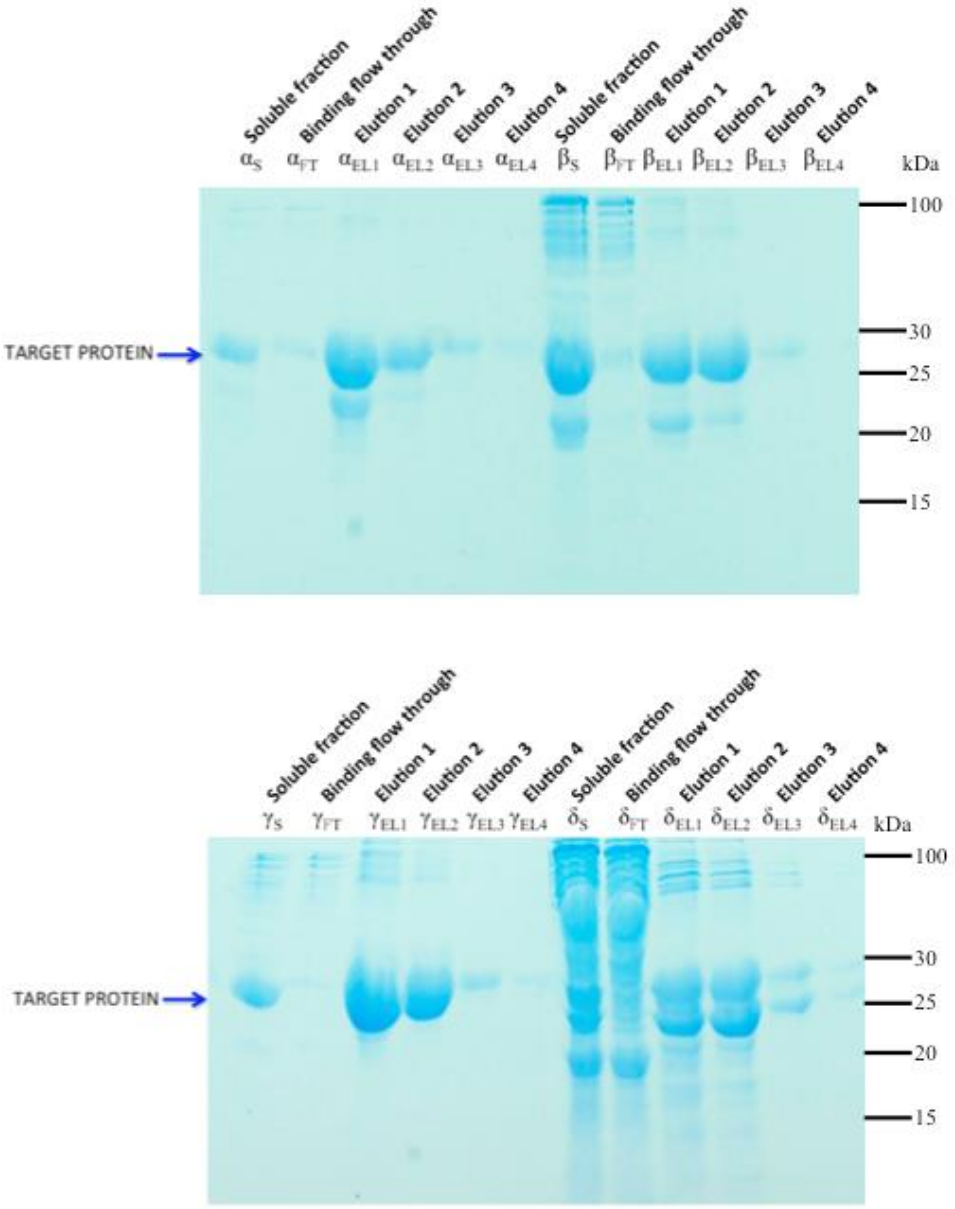


Figure 4.7 | Analysis of large-scale purified recombinant AtPCP-B proteins by SDS-PAGE. The SDS-PAGE gels (18%) with soluble fraction (α_S, β_S, γ_S, δ_S), binding flow through (α_{FT}, β_{FT}, γ_{FT}, δ_{FT}) and eluted samples ($\alpha_{EL1\text{-}4}$, $\beta_{EL1\text{-}4}$, $\gamma_{EL1\text{-}4}$, $\delta_{EL1\text{-}4}$) of large-scale purified protein samples by His-trap FF.

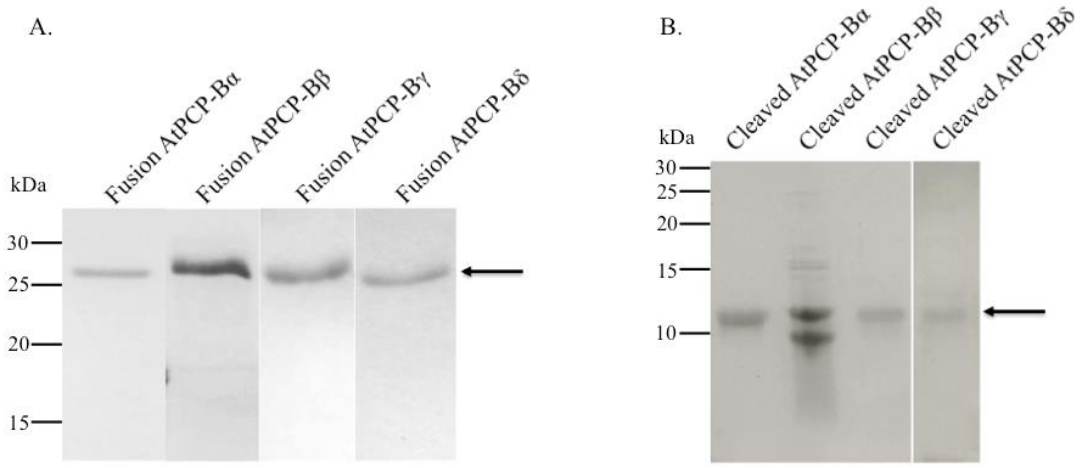


Figure 4.8 | SDS-PAGE (18%) of purified AtPCP-B α, β, γ and δ fusion proteins (A) and mature proteins after enterokinase cleavage (B). Fusion proteins were generated using the pET32a vector and expression was in in *E.coli* (*Origami 2* (DE3)). Purification was carried out with His-trap FF column. The purification of cleaved mature proteins was carried out by reverse phase chromatography. The arrow indicates fusion protein and mature protein bands.

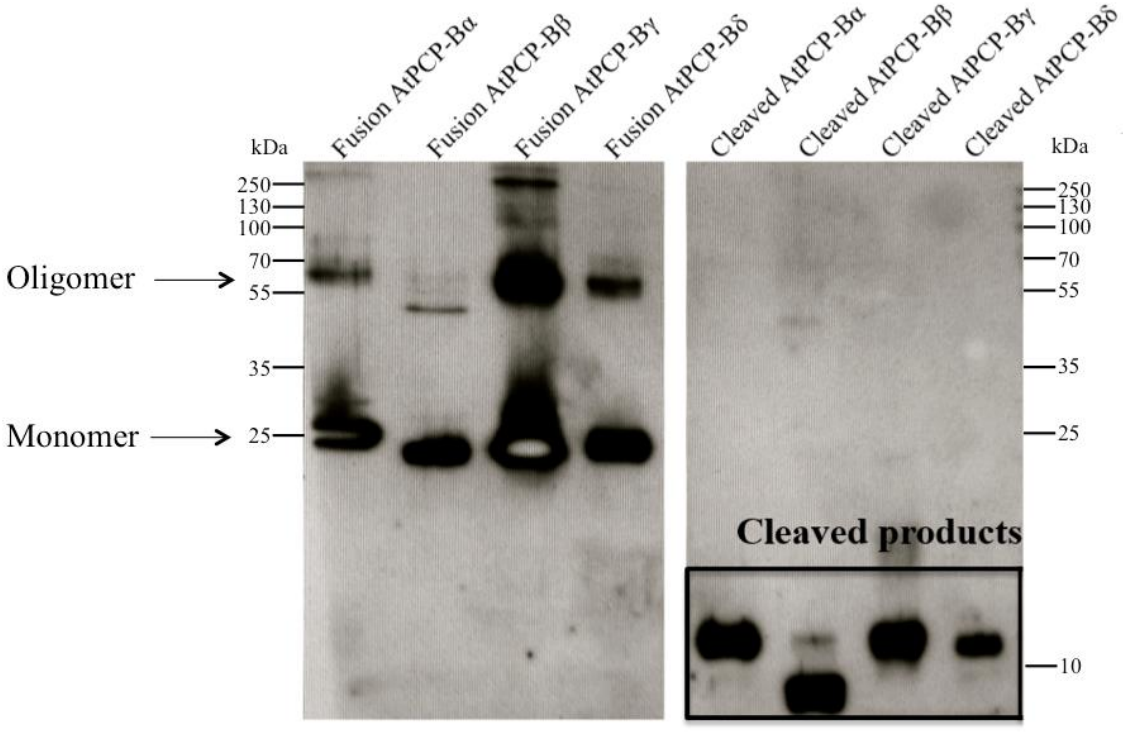


Figure 4.9 | Western blot of purified heterologously expressed AtPCP-B recombinant proteins and cleaved AtPCP-B mature proteins (His-PCP-Bs). The products were diluted 100 times before sample loading. Arrows indicate the bands of recombinant proteins (monomer) and possible oligomers. A box highlights bands of His-PCP-Bs (cleaved products). Primary antibody: 6x His Rabbit Polyclonal antibody; Secondary antibody: goat anti-rabbit IgG-HRP.

4.2.4 Detecting the interaction of AtPCP-Bγ with putative stigmatic protein targets

To identify potential stigmatic targets of AtPCP-Bs, three protein-protein interaction assays were carried out in this study, including far-western blotting, protein pull-down and protein crosslinking assays. Based on the loss-of-function study reported in Chapter 3, AtPCP-Bγ showed the most significant pollen hydration phenotype (Wang L *et al*., 2016). Thus, we hypothesised that AtPCP-Bγ could potentially have the highest level of affinity with a target protein and thus this protein was chosen as 'bait' in the interaction assays.

Far-western blotting is a method derived from standard western blotting (WB) to investigate protein-protein interactions (Wu *et al.*, 2007). Arabidopsis stigmatic proteins (prey proteins) were first separated by SDS-PAGE and transferred onto a PVDF membrane as for standard WB (Figure 4.10a). The membrane was then probed with the purified His-tagged AtPCP-Bγ protein (bait protein) and detected using an anti-His antibody (Figure 4.10b). Protein-protein interactions or formation of protein complexes generally requires correct protein folding and conformation. In case protein-protein binding was adversely affected by the process of protein denaturation, a stigmatic total cell protein sample without β-ME treatment (native) was also used in this work. Results from the far-western blotting indicated that AtPCP-Bγ bound to two protein bands having estimated sizes of 80kDa and 130 kDa (Figure 4.10b). The band from the denatured sample was likely to be the consequence of the binding between AtPCP-Bγ with the most abundant stigmatic protein, whereas the band from the native sample was probably the product of the binding of AtPCP-Bγ with the second most abundant stigmatic protein on the gel (Figure 4.10b). The differences of the blotting patterns from the denatured and native samples indicated that the β-ME treatment did affect the protein interaction by altering the property of protein folding or conformation. Although this result demonstrated a possible interaction of AtPCP-Bγ with two stigmatic proteins, the abundance of the stigmatic proteins indicated that these interactions were likely to be non-specific.

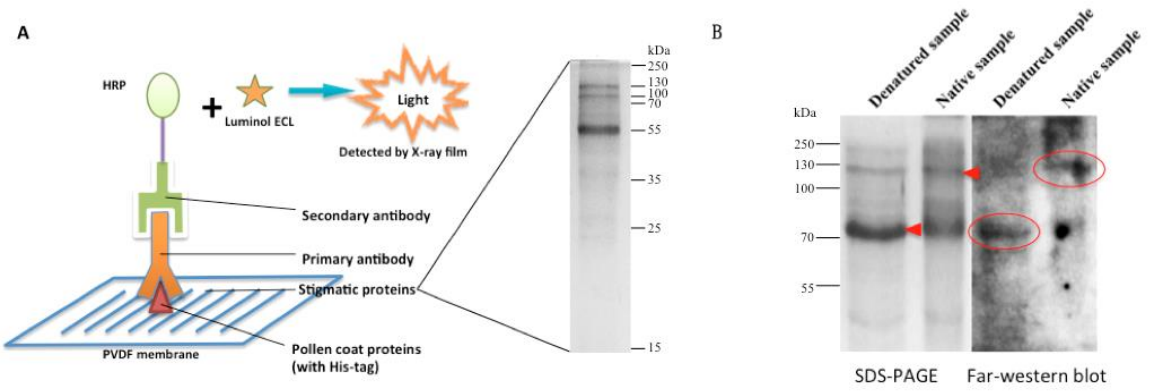


Figure 4.10 | Far-western blotting of the AtPCP-Bγ-His fusion protein with a total Arabidopsis stigmatic protein extract. A. The diagram shows the basis of far-western blotting. B. SDS-PAGE gel with isolated total stigmatic proteins with (reduced) and without (native) β-ME treatment, and the corresponding x-ray film after interaction with the AtPCP-Bγ-His ligand. Bands indicating a potential AtPCP-Bγ binding are highlighted.

As AtPCP-Bs may interact with their stigmatic targets transiently without forming a stable complex or that the interactions involving heterologously expressed protein *in vitro* may be weak, protein crosslinking was utilised to capture transient or low-affinity interactions between PCP-B ligands and their potential stigmatic targets. Glutaraldehyde is a widely used protein crosslinker and is useful for stabilising structural intra and intermolecular interactions by reacting with amine groups of lysine residues in proteins (Habeeb & Hiramoto, 1968). To achieve the greatest similarity with *in vivo* conditions for this assay, the crosslinking reaction was carried out under mild conditions of pH and temperature (close to that in living plant tissue) to preserve the native structures and activity of the proteins. Cleaved mature AtPCP-Bγ-His with N-terminal tags removed was used to eliminate the effect of large N-terminal tags in the reaction and a total stigmatic protein extract was used as 'prey'. Interacting complexes were then captured by virtue of the AtPCP-Bγ His-tag using a nickel affinity IMAC gel. In order to identify stigmatic proteins that bound non-specifically to the nickel affinity IMAC gel, an additional control sample was subjected to purification where the AtPCP-Bγ solution was replaced with the same volume of reaction buffer. Crosslinked complexes were then eluted from the nickel affinity IMAC gel and analysed by using LC-MS/MS. The statistically validated data set identified three proteins as components of the crosslinked product (Table 4.4), of which the most abundant protein was AtPCP-Bγ (Accession: A8MR88), indicating that the AtPCP-Bγ was efficiently isolated by the column. No peptides were detected in the control sample, suggesting that the stigmatic proteins identified in the experimental sample were indeed

crosslinked to AtPCP-Bγ. However, compared with the mean frequency of lysine residues (5.8%) in proteins (Trinquier & Sanejouand, 1998), the amino acid sequences of the interacting stigmatic proteins revealed a high lysine content: 43 lysine residues (8.4%) in the heat shock cognate protein 70-1 (Hsc 70-1) and 83 (9.2%) in Histone-lysine N-methyltransferase (CLF). Due to this high lysine content that facilitates the binding of molecules to glutaraldehyde, the validity of the interactions of these two proteins with AtPCP-Bγ must be treated with caution.

Table 4.4 | Proteins* detected by mass spectrometry in crosslinking assay with AtPCP-Bγ and a total stigmatic protein extract.

Accession	Description	Score[a]	Coverage[b]	Proteins[c]	Unique Peptides[d]	Peptides[e]	PSMs[f]	Area[g]	AAs	MW/kDa	pI[h]
A8MR88	EMBRYO SURROUNDING FACTOR 1-like protein 8	9.31	30.26	1	3	3	4	3.153E5	76	8.3	7.47
F4KCE5	Heat shock cognate protein 70-1	4.76	3.07	10	1	1	1	7.083E4	521	57.2	5.10
P93831	Histone-lysine N-methyltransferase CLF	0.00	2.11	1	1	1	1		902	100.3	8.60

a. The sum of the scores of the individual peptides
b. The percentage of the protein sequence covered by identified peptides
c. The number of identified proteins in the protein group of a master protein
d. The number of peptide sequences unique to a protein group
e. The number of distinct peptide sequences in the protein group
f. The number of peptide-spectrum match (PSM) for the protein, including those redundantly identified
g. The average area of the three unique peptides with the largest peak area
h. Theoretically calculated isoelectric point, which is the pH at which a particular molecule carries no net electrical charge
* The PSMs were statistically validated to avoid false positives by using the False discovery rate (FDR)-controlling procedure (for details see 2.9.4).

A further approach was taken that utilised a pull-down assay to detect potential stigmatic targets of AtPCP-Bγ. The histidine-tagged mature AtPCP-Bγ was first immobilised onto IMAC nickel affinity gel and acted as 'bait'. A stigmatic microsomal membrane fraction (prey) was loaded onto the prepared gel to interact with the 'bait'. Binding products were eluted from the nickel affinity gel and analysed by using LC-MS/MS as described for the protein crosslinking assay above. The MS results revealed a number of possible stigmatic AtPCP-Bγ binding partners whereas only one protein was detected from the control sample (which lacked bait protein), demonstrating that stigmatic membrane proteins had little non-specific affinity with the nickel affinity gel (Table 4.5). As was found for the protein crosslinking data set, AtPCP-Bγ was the most abundant protein in the purified pull-down product, confirming the high efficiency and specificity of the affinity between the His-tagged protein and the nickel affinity gel. However, several of the more abundant proteins identified in the data set from the pull-down assay included ribosomal proteins and dead-box proteins (Table 4.5). These proteins, along with the heat shock (cognate) protein that was detected in the protein crosslinking experiment, have previously been identified as common non-specific binding proteins in proteomic analyses and protein pull-down assays (Trinkle-Mulcahy *et al.*, 2008). Thus, although a control sample that lacked column-immobilised AtPCP-Bγ gave a relative 'clean' elution, the possibility of non-specific binding of stigmatic membrane proteins to the column cannot be ruled out. Despite the presence of these likely non-specific interactors, several uncharacterized predicted proteins were also detected, which could also be considered as candidate stigmatic targets of AtPCP-Bγ. Interestingly, when compared the pull-down data set with the proteomes of stigmatic membrane proteins and total cell protein extract as part of a side study in this project (Table 4.5, Appendix 2.8, 2.9), a stigmatic membrane protein, C2 domain-containing protein was detected from both the pull-down product and stigmatic membrane fraction. As denoted in UniProtKB (http://www.uniprot.org/uniprot/Q93XX4), C2 domain-containing protein At1g53590 possesses two transmembrane features due to the presence of α-helical transmembrane regions.

Table 4.5 | Statistically validated hits from mass spectrometry analysis of proteins obtained from the AtPCP-Bγ pull-down assay. a. Data from the experimental sample; Proteins found in the proteome of the Arabidopsis stigmatic membrane fraction (Appendix 2.8) were labelled with empty circles, the ones found in proteome of Arabidopsis stigmatic total cell protein extract (Appendix 2.9) were labelled with filled circles. b. Data from the control with no bait protein AtPCP-Bγ in the pull-down reaction.

a.

Accession		Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW/kDa	pI
A8MR88		EMBRYO SURROUNDING FACTOR 1-like protein 8	2451.41	59.21	1	8	8	1204	3.770E8	76	8.3	7.47
Q9SGA6		40S ribosomal protein S19-1	12.08	18.18	1	1	3	6	5.175E5	143	15.8	10.08
Q9FNP8	●○	40S ribosomal protein S19-3	11.81	18.18	2	1	3	6	4.699E5	143	15.7	10.21
F4JLA9		Cruciferin 3	10.14	15.89	2	3	3	3		453	50.0	7.14
F4K8S2		12S seed storage protein CRU1	5.89	7.72	2	1	1	2	3.277E5	285	31.6	8.81
P34788	●○	40S ribosomal protein S18	5.18	7.24	1	1	1	2	7.667E5	152	17.5	10.54
Q94AU7		Gamma carbonic anhydrase 3, mitochondrial	3.78	7.36	1	1	1	1		258	27.8	7.27
A8MS83		60S ribosomal protein L23a-2	3.69	11.49	3	1	1	1		148	16.7	10.17
Q9S7Y7	●	Alpha-xylosidase 1	3.62	1.09	1	1	1	2	4.746E5	915	102.3	6.77
Q9C6B3		Gamma carbonic anhydrase 2, mitochondrial	2.56	10.07	1	1	1	1		278	30.0	7.24
Q2V2Y5		AT hook motif-containing protein	2.43	4.59	3	1	1	1	1.636E6	283	31.2	8.27
A0A097PSE1		S-adenosyl-L-methionine-dependent methyltransferase superfamily protein (Fragment)	2.26	5.42	1	1	1	1		332	37.1	8.62
Q0WPG1		Exostosin family protein	2.21	2.91	1	1	1	1	8.936E5	654	73.5	9.51
Q85B88		Ribulose 1,5-bisphosphate carboxylase/oxygenase large chain (Fragment)	2.15	2.56	4	1	1	1	2.618E5	430	47.5	6.58
Q93XX4	○	C2 domain-containing protein At1g53590	0.00	2.40	1	1	1	1	3.174E4	751	84.8	5.99
Q8GVE8		Phosphoenolpyruvate carboxylase 4	0.00	2.33	1	1	1	1		1032	116.5	7.12
Q93ZT6		Eukaryotic translation initiation factor isoform 4G-1	0.00	2.18	1	1	1	14		780	85.5	8.24
F4IUG9		Myosin-13	0.00	0.74	1	1	1	1		1493	169.0	7.43
Q9SFW8		Protein NLP5	0.00	0.99	1	1	1	2		808	90.6	6.19
Q9LUW6		DEAD-box ATP-dependent RNA helicase 9	0.00	3.93	1	1	1	1	3.243E5	610	63.6	9.41
Q9SR73		40S ribosomal protein S28-1	0.00	18.75	1	1	1	1	1.128E5	64	7.4	10.83
O82746		Putative uncharacterized protein AT4g22980	0.00	1.79	1	1	1	1	1.197E6	559	64.0	6.73
Q94BZ0		At1g50450/F11F12_20	0.00	3.04	1	1	1	1		428	46.5	8.09
Q9XI31		AT1G15380 protein	0.00	5.75	1	1	1	4	3.212E6	174	19.6	5.91
F4KE63		ATP binding/valine-tRNA ligase/aminoacyl-tRNA ligase	0.00	1.23	1	1	1	1	3.154E5	974	110.6	6.37
F4J2K4		Kinesin-like protein	0.00	0.39	1	1	1	1	2.898E5	1273	144.7	5.40
F4KEW8		Protein NETWORKED 4A	0.00	1.61	1	1	1	1		558	64.1	5.22
Q9M067		ATM-like protein (Fragment)	0.00	0.81	5	1	1	1	3.018E5	2089	236.7	6.71
Q9LHD6		Genomic DNA, chromosome 3, P1 clone: MZF16	0.00	7.58	1	1	1	1	5.844E6	211	24.4	8.70
Q9XI58		Cytochrome b-c1 complex subunit 6	0.00	20.79	1	1	1	2		101	11.3	6.73

b.

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW/kDa	pI
Q9SV13	Sulfate transporter 3.1	0.00	5.17	1	1	1	1	1.363E7	658	72.7	8.91

4.3 Discussion

It has long been a challenge to identify the unknown interacting targets of ligands or receptors. The widely used classic yeast-two-hybrid approaches are known to be problematic for detecting extracellular interactions and thus this approach was not used in this work (reviewed in Bruckner *et al.*, 2009). In this chapter, three protein-protein interaction methods were utilised to detect any possible binding between AtPCP-B ligands and stigmatic 'receptors'. Due to time constraints and to avoid unnecessary experimental complexity, AtPCP-Bγ was used to generate these preliminary results due to the fact that mutation of its respective gene generated a strong phenotype when compared with other AtPCP-Bs. Although only AtPCP-Bγ was used for protein-protein interaction studies, the successful heterologous expression of all four AtPCP-Bs will likely facilitate future work on the interactions that occur at the pollen-stigma interface.

4.3.1 Cloning of the *AtPCP-B* genes

The *AtPCP-B* sequence data obtained from the cloned cDNAs for PCP-B α, γ and δ was as predicted and thus provided independent confirmation of the publicly available data for Arabidopsis (http://www.arabidopsis.org). However for *AtPCP-Bβ* (which resides in the annotated gene *At2g29790*), alternative splicing of the *At2g29790* gene was confirmed (4.2.2). The mature PCP-Bβ encoding region used for cloning only starts in the last exon of *At2g29790*, which is one of the shared exons between *At2g29790* and *AtPCP-Bβ.* Thus, although the splicing of *At2g29790* was different from previous predictions (Figure 4.2), there was no alternative to the cloning of cDNA sequence that encodes mature PCP-Bβ in this work.

4.3.2 Protein expression and purification

The correct formation of disulphide bonds is crucial for producing soluble and active proteins. In a previous study, a Brassica *SCR/CRP* gene was cloned into the pET-32a vector and expressed in the Origami^{TM} cell line which facilitated the formation of disulphide binds (Kemp & Doughty, 2007). Interestingly, multiple SCR isoforms were identified by reverse-phase high-performance liquid chromatography (HPLC) from the purified expression products (Kemp & Doughty, 2007). Some of these isoforms were likely caused by the formation of incorrect disulphide bonds in the SCR polypeptide. The

refolding of insoluble protein or bacterial inclusion bodies may have caused incorrect disulphide bond formation that could lead to the inactivity of protein products. Thus, in this project, only the soluble protein fraction prepared from the expression cell culture was used for protein purification. The optimisation of AtPCP-B protein expression conditions was generally successful for increasing the proportion of soluble protein. However in the case of AtPCP-Bβ, the proportion of soluble protein following expression in *E. coli* remained relatively low compared to the other three AtPCP-Bs. Despite proportionately low yields, by increasing the volume of expression cell culture larger quantities of AtPCP-Bβ could be readily obtained.

For the purification of His-tagged proteins with a prepacked IMAC nickel affinity column, the binding specificity largely depends on the column flow rate and the buffer environment. The addition of imidazole as a competitive agent in the binding buffer improves purity of His-tagged proteins (Bornhorst & Falke, 2000). To minimize non-specific binding, the optimal imidazole concentrations of buffers needed to be determined according to different proteins. Thus, the flow rate and buffer environment required for large-scale purification of PCP-B proteins will need further optimisation to obtain high purity proteins.

4.3.3 Verification of recombinant protein expression and quantification

Comparing the protein profiles of the purified recombinant AtPCP-Bs by SDS-PAGE and western blotting (utilising a 6x His rabbit polyclonal antibody), it was noted that higher molecular weight species were identified that cross-reacted with the antibody that could not be detected on the Coomassie stained gels (Figure 4.8). Thus, it seems probable that a small proportion of the recombinant AtPCP-B protein forms homo-oligomers (dimers, trimers etc.) that account for these bands. This result indicates that the anti-His antibody and ECL detection system is highly sensitive and can detect very low levels of His-tagged AtPCP-B proteins. Importantly this sensitivity permitted these His-tagged proteins to be utilised in far-western blotting approaches (described in this chapter) to detect protein-protein interactions between the AtPCP-Bs and putative stigmatic target proteins.

4.3.4 Identification of potential stigmatic targets for AtPCP-B ligands

Evidence from the functional study of AtPCP-Bs (Chapter 3, Wang L *et al*., 2016) suggested that AtPCP-Bs are likely to act as ligands for stigmatic targets by directly or

indirectly activating regulators of water transfer through the stigmatic papillar cell plasma membrane. In addition AtPCP-Bs were found to be closely related to ESF1 which is proposed to act as an extracellular signal to shape embryo development in Arabidopsis (Costa *et al.*, 2014). Although a large number of methods have been developed for the study of protein-protein interactions, this project had to start with an initial screen of binding targets amongst all extractable stigmatic proteins as no potential 'receptors' were known. Thus, three *in vitro* techniques were selected that utilise affinity-tagged proteins to capture putative AtPCP-B binding partners and these included far-western blotting, protein crosslinking and protein pull-down assays.

By considering the interaction of SCR and SRK, a well-studied protein interaction example between a pollen-borne ligand and a stigmatic target, we proposed that the stigmatic target(s) of AtPCP-Bs may be present on the outer surface of papillar plasma membrane. Far-western blotting was first carried out using total stigmatic proteins rather than focussing on membrane proteins due to the relative ease of obtaining such a protein sample. Although the initial blotting results showed a significant degree of non-specific binding, the different blotting patterns obtained between denatured and native stigmatic samples indicated the effect of protein folding to their interaction with AtPCP-B ligand. The data obtained using this approach was considered unreliable and thus the study moved to crosslinking and pull-down approaches to identify putative AtPCP-B targets.

The protein crosslinking and pull-down assays followed by mass spectrometry analyses revealed a number of proteins that could potentially interact with AtPCP-Bγ. However ribosomal proteins were abundant in the data set and these were probably the result of non-specific binding. In theory, if PCP-Bs are interacting with proteins present in the plasma membrane of stigmatic papilla cells then protein targets identified in the pull-down assay should be present in the proteome of the microsomal membrane fraction. However, only a small number of proteins identified from the pull-down assay were found to be present in the proteome of the stigmatic membrane (Appendix 2.8), while some stigmatic proteins detected from pull-down assay were showed in proteomes of both stigmatic cell extract (Appendix 2.9) and stigmatic membrane (Appendix 2.8). Indeed most putative AtPCP-B target proteins detected from pull-down assay were found neither in the total protein extract from stigmatic cells nor in the proteome of stigmatic membrane (Table 4.5, Appendix 2.8, 2.9). There are two potential explanations for these results; i) the proteomic

analysis of the stigmatic microsomal membrane fraction (Appendix 2.8) did not detect proteins with low abundance due to the scarcity of sample material, ii) the microsomal membrane fraction (Appendix 2.8) was contaminated by the total protein fraction (Appendix 2.9) during extraction thus some of the total cell proteins were detected from the elution of pull-down assay. Thus, the lack of coherence between the data sets generated by pull-down assay and crosslinking suggests that much of the data obtained represents false positives. In fact, by comparing methods used for large-scale protein-protein interaction analyses, it has been estimated that as many as 30-60% of interactions detected by high-through-put studies including the yeast-two-hybrid system, affinity-based approaches or *in silico* methods represent false positives (von Mering *et al.*, 2002). Thus, it is crucially important to validate protein-protein interaction by using multiple methods.

Interestingly, a C2 domain-containing protein At1g53590 was found to be a potential AtPCP-Bγ interactor amongst stigmatic membrane proteins derived from the pull-down assay. The C2 domain is a Ca^{2+}-binding motif that has been found in eukaryotic signalling proteins that mediate multiple intracellular activities such as vesicular transport, membrane fusion, and lipid modification (reviewed in Nalefski & Falke, 1996). As an important regulator in angiosperms, calcium (Ca^{2+}) has been found to have a broad range of signalling functions such as attraction of pollen tube growth with accumulated Ca^{2+}, cellular communication and long /short- distance signalling (reviewed in (Ge *et al.*, 2007). During plant reproduction, calcium (Ca^{2+}) has been found to have a multiplicity of roles including being essential for pollen germination, pollen tube elongation and guidance, pollen tube discharge and gamete fusion (Franklin-Tong *et al.*, 2002; Iwano *et al.*, 2004; Sprunck *et al.*, 2012; Steinhorst & Kudla, 2013). Calcium dynamics have been observed *in vivo* immediately post-pollination in hydrating pollen and the papillar cell close to the pollen-stigma contacting site, as well as in the growing tip of pollen tube (Iwano *et al.*, 2004). In plants, Ca^{2+}-sensing proteins that contain the C2 domain have been found to be involved in the regulation of ABA signalling (Rodriguez *et al.*, 2014) as well as defence responses to abiotic and biotic stress (de Silva *et al.*, 2011; Zhang *et al.*, 2013). Importantly a functional connection between a leucine-rich repeat receptor-like kinase (LRR-RLK) STRUBBELIG (SUB) and QUIRKY (QKY), a membrane protein that contains four C2 domains, has recently been described (Vaddepalli *et al.*, 2014). Based on this discovery it is tempting to propose a model that AtPCP-Bs indirectly regulate water transport through the papillar cell membrane by interacting with and modifying a C2 domain-containing

protein (At1g53590), which may mediate vesicular trafficking through calcium dynamics in the contacting papillar cell. Although it has been predicted that the C2 domain-containing protein At1g53590 possesses two α-helical transmembrane motifs (http://www.uniprot.org/uniprot/Q93XX4), no empirial evidence was found to confirm this feature. Further protein-protein interaction assays, papilla cell transcriptome data and investigation on pollination phenotype in T-DNA knockout lines should be utilised for validating this candidate interactor.

Although the detection of heat shock cognate protein 70-1 (Hsc 70-1) and ribosomal proteins by LC-MS/MS following the protein-protein interaction analyses was likely to be a consequence of non-specific binding, the possibility cannot be ruled out that these proteins might act as important indirect regulators of the pollen-stigma interaction and pollen hydration. Hsc 70-1 is a highly homologous chaperone to the heat shock protein 70 (Hsp70) family. In plants, Hsp70 is an important molecular chaperone involved in proper protein folding and microbial pathogenesis (reviewed in Park & Seo, 2015). Interestingly, members of the Hsp70 family have also been found to be involved in processes related to plant reproduction, which suggests there is a possibility that Hsc 70-1 could act as an indirect regulator of pollen hydration. For instance, immunoglobulin-binding proteins (BiPs), the molecular chaperones in the heat shock protein 70 (Hsp 70) family, were found to be expressed in pollen and pollen tubes of *Arabidopsis thaliana* (Maruyama *et al.*, 2014). Loss-of-function studies for *BiPs* revealed defects in male gametogenesis and pollen tube growth, effects that may have been caused by decreased efficiency of protein translocation, protein folding and quality control of proteins in the endoplasmic reticulum (ER) (Maruyama *et al.*, 2014). With regards to potential roles for ribosomal proteins and molecular chaperone, a group of cysteine-rich proteins (CRPs), the nodule-specific cysteine-rich (NCR) proteins were found to be exclusively expressed in the rhizobium-infected symbiotic plant cells and provoked terminal bacterial differentiation (Van de Velde *et al.*, 2010). A study focused on NCR247 discovered that this CRP binds to the chaperone GroEL and forms a complex with ribosomal proteins, which together may mediate the differentiation processes of bacteroids in infected roots (Farkas *et al.*, 2014).

Although it has not been possible to identify with any degree of confidence the interacting target(s) of AtPCP-Bs during this project, previous studies on protein signalling pathways mediated by CRPs, the regulations of plant reproduction and the interaction of plant cells

and the biotic factors in its environment inspired hypothetical candidates. For instance, Receptor kinases / Receptor-like kinases (RLKs) have been identified as the binding targets of CRPs in reproductive signalling. The male determinant of self-incompatibility (SI), *S*-locus cysteine-rich protein (SCR) binds to the female determinant S-receptor kinase (SRK), a membrane-spanning serine/threonine kinase (Takayama *et al.*, 2001). Two receptor-like kinases, LIP1 and LIP2, localised to the pollen tube tip cytoplasmic region, have been identified as essential receptor complex components in pollen tube guidance signalling (Liu *et al.*, 2013). In addition it has recently been shown that the heterodimer MDIS1-MIK acts as the receptor for AtLURE1 peptides that are expressed in synergid cells and act as attractants for pollen tube guidance to the ovule micropyle (Takeuchi & Higashiyama, 2012; Wang T *et al.*, 2016).

Nevertheless, the stigmatic targets of AtPCP-B ligands are currently unknown, though several candidates have arisen from the current data sets of protein-protein interaction studies in this project. It is quite possible, based on our current knowledge of receptor ligand interactions, that perception of the pollen PCP-Bs by the stigmatic target(s) may be far more complicated than we initially expected. A recent study of the interaction between AtLURE1 and its heterodimer receptor MDIS1-MIK revealed an additive effect of pollen tube attraction in the double gene mutant mdis1/2, while AtLURE1 binds to MDIS1 but not MDIS2 (Wang T *et al.*, 2016). Since the loss-of-function studies of *AtPCP-Bs* showed a similar complex combinatorial phenotype (i.e. when combining the *pcp-bα* gene mutation to the *pcp-bβ/γ* double mutant, a striking enhancement of the hydration defect was observed - see chapter 3, Wang L *et al*., 2016), it is possible that a similar scenario might apply to AtPCP-Bs; that is there are multiple stigmatic targets that could potentially interact with AtPCP-Bs and more than one ligand may interact with the same target.

The heterologous expression of AtPCP-Bs facilitated the search for their female receptors, although the study may have been hampered by the redundancy of *AtPCP-B* genes and their possible transient interaction with stigmatic targets. Preliminary results from the protein-protein interaction assays did identify candidate AtPCP-B binding targets and provided experience in the use of multiple methods that can be applied to further future studies in the hunt for female receptors of the AtPCP-B ligands.

Chapter 5 Phylogenetic and evolutionary analysis of PCP-B like proteins

5.1 Introduction

The 'B' class of pollen coat proteins have been discovered to play a crucial role in regulating early stages of the pollen-stigma interaction (chapter 3, Wang L *et al*., 2016). The loss-of-function studies of *PCP-B* genes in *Arabidopsis thaliana* suggested that these proteins are acting as regulatory factors for basal compatibility. Interestingly, another three homologous *PCP-B* genes, *Embryo surrounding factor 1.1-3* (*ESF1.1-3*), were recently identified as important regulators of early embryo development (Costa *et al.*, 2014). These discoveries highlight the fact that the *PCP-B* gene family has undergone functional diversification. Amongst the angiosperms, their rapid diversification and speciation that took place over a relatively short evolutionary period (~ 120 million years) is held to have been largely driven by the evolution of reproductive barriers. Plant species are isolated by multiple reproductive barriers, including pre-pollination barriers that limit transfer of heterospecific pollen to stigmas, post-pollination/pre-zygotic barriers that prevent non-conspecific pollen from fertilising eggs and post-zygotic barriers that causes hybrid inviability, offspring sterility, and hybrid breakdown (reviewed in Rieseberg & Willis, 2007).

In recent years, many genes have been identified as 'speciation genes', which contribute to the formation of reproductive barriers by reducing the amount of gene flow between populations (Reviewed in Rieseberg & Blackman, 2010). Though significant progress has been made in understanding the molecular and genetic basis of post-zygotic reproductive isolation between species (Reviewed in Rieseberg & Blackman, 2010), much less is known about pre-zygotic barriers that act in the early stages of plant reproduction. As mentioned in Chapter 1 (see section 1.1.3), early-acting reproductive barriers such as pre-pollination and post-pollination prezygotic barriers, contribute more than the postzygotic barriers to speciation. Mutation of genes that contribute to the pre-pollination prezygotic barrier may alter the floral pigments, flowering time and leads to shifts in mating strategy from outcrossing to selfing (Chen *et al*., 2007; Schwartz *et al*., 2009). For example, the mutation of a structural gene, FLAVONOID-3'-HYDROXYLASE (F3'H) affects the anthocyanin pathway, which results in the transition of floral colour from blue/ purple for bee

pollination to red for hummingbird pollination (Des Marais & Rausher, 2010). For genes which contribute to the post-pollination prezygotic reproductive barriers, most examples involve mutations of genes at the *S*-locus such as those encoding *S*-RNase that regulate self-incompatibility (SI) in Solanaceae and SCR/SRK proteins that regulate SI in Brasicaceae (see Chapter 1, 1.1.2). The recently identified synergid cell-derived LUREs were discovered to be pollen tube attractants by interacting with a male pollen-tube receptor (Takeuchi & Higashiyama, 2012; Wang T *et al.*, 2016). The heterologous expression of AtLURE1 in the synergid cells of *Torenia fournieri* was sufficient to guide Arabidopsis pollen tubes into the *Torenia fournieri* embryo sac (Takeuchi & Higashiyama, 2012). Genome-wide comparative sequencing studies in multiple species show that genes that mediate sexual reproductive processes evolve more rapidly than other genes (reviewed in Clark *et al.*, 2006).

There is a growing body of evidence that the rapid divergence of reproductive proteins is likely to be the result of adaptive selection (Swanson & Vacquier, 2002). Previously, there have been very few reproductive proteins in angiosperms reported to be under positive selection (Clark *et al.*, 2006), examples include the sporophytic SI proteins SCR, SRK, and SLG (Sato *et al.*, 2002; Takebayashi *et al.*, 2003) and gametophytic SI proteins *S*-RNase and SLF (Takebayashi *et al.*, 2003; Ikeda *et al.*, 2004). By interacting with stigmatic targets, PCP-Bs may also be involved in the formation of a post-pollen-deposition, pre-zygotic barrier. Within this context, given that the *PCP-B* genes under study here function to regulate one of the earliest steps of plant reproduction in the Brassicaeae, it was considered that an analysis of the molecular evolution of the *PCP-B*-like (*PCPBL*) gene family would be valuable. We hypothesised that the PCP-B pollen ligands are likely undergoing rapidly diversification and evolving under positive selection to adapt to the evolution of their interacting receptors. In this chapter, a phylogenetic analysis and molecular evolutionary study of genes encoding the PCP-B-like protein family is reported. This study provides evidence that rapid gene birth-and death along with adaptive selection is a feature of *PCP-B* gene family evolution. This work might facilitate the correlation of sequence divergence with functional diversification of pollen coat proteins after gene duplication and reveal functionally important regions of these molecules.

5.2 Results

5.2.1 *PCP-B*-like genes are present in multiple angiosperm families

Previously, 14 genes have been identified in the *Arabidopsis thaliana* genome encoding PCP-B-like (PCPBL) sequences (Silverstein *et al.*, 2007). The unique eight-cysteine residue pattern enabled us to identify other *PCP-B* homologues in plants. In addition to these 14 PCPBL protein sequences in *Arabidopsis thaliana*, 280 genes encoding PCPBL sequences were identified from seven angiosperm families by using tBLASTn searches against all currently available sequenced plant genomes (Table 5.1). Two families, the Brassicaceae and Poaceae, had the most genomes containing PCPBL-encoding genes. The other five families, Malvaceae, Nelumbonaceae, Pedaliaceae, Solanaceae and Phrymaceae, contained only one or two genomes with PCPBL-encoding genes. These observations were probably due to the relative higher coverage of available genome sequencing data in the Brassicaceae and Poaceae compared to the other five families, though not all available genomes contain *PCPBLs.* Interestingly, in addition to the 14 previously identified *PCPBL* genes in Arabidopsis, another *PCP-B* homologous gene, *At1g10705*, although designated as a putative pseudogene, possessed an amino acid sequence that resembles a PCPBL sequence. Additionally, a total of 36 PCPBL protein sequences were identified from other ecotypes of *Arabidopsis thaliana* (not including Col-0), including 4 in Bur-0, 5 in C24, 6 in Can-0, 5 in Ct-1, 6 in E di-0, 6 in Kro-0 and 4 in Ler-1. Currently completion percentages of ecotype genomes include Bur-0, 80.64%; C24, 81.95%; Can-0, 98.07%; Ct-1, 98.29%; Edi-0, 98.50%; Kro-0, 80.80% and Ler-1, 80.44% (divided by Col-0 119,667,750bp). However, no obvious correlation was observed between the numbers of homologous PCP-B genes and predicted genome size. It is important to note that although fewer gene copies were identified in some taxa than others, which could have been a result of the complex evolutionary history of *PCPBL* genes, it is very likely that the availability and degree of completion of plant genome sequences also impacted on the identification of *PCPBL* genes. Thus the numbers reported here are likely to be an underestimate of the true numbers of *PCPBL* genes in several of the taxa studied.

To investigate the phylogenetic relationship and molecular evolution of the *PCP-B* and *PCPBL* genes in plants, homologous genes were retrieved only from species with completed and assembled genome sequences, representing mainly taxa of the Brassicaceae

and Poaceae (Table 5.1). A functional cysteine-rich protein consists of an N-terminal signal peptide and a C-terminal cysteine-rich region, which are usually encoded by genes with two exons and one intron. To avoid putative pseudogenes, only the ones that encode full-length PCPBL proteins were selected for further analysis. The final working data set contained 134 *PCP-B* homologous genes from 15 species that included taxa from the Brassicaceae, Malvaceae and Poaceae (Table 5.2). The sizes of the open reading frames (ORFs) encoding mature proteins (without predicted signal peptide) ranged from 120 to 327 base pairs (bp). A reference gene family, *Ubiquitin-like 5*, was selected for the evolutionary comparative analysis (Table 5.2) because of the similarity of gene size with the *PCPBL*s and their non-reproductive function.

Table 5.1 | Numbers of PCP-B-like (PCPBL) sequences identified in species from seven angiosperm families.

Family	Species	Numbers of PCPBL sequences in species	Numbers of PCPBL sequences in family
Brassicaceae	*Arabidopsis thaliana*	15(Col-0)+36* (ecotypes)	160
	Arabidopsis lyrata	14	
	Capsella rubella	10	
	Capsella grandiflora	12	
	Capsella orientalis	13	
	Neslia paniculata	3	
	Camelina sativa	4	
	Leavenworthia alabamica	4	
	Boechera stricta	15	
	Arabis alpina	3	
	Brassica oleracea	10	
	Brassica rapa	8	
	Brassica napus	16	
	Raphanus raphanistrum	6	
	Raphanus sativus	9	
	Sisymbrium irio	3	
	Eutrema salsugineum	7	
	Tarenaya hassleriana	8	
Malvaceae	*Gossypium raimondii*	5	6
	Gossypium arboreum	1	
Poaceae	*Oryza sativa*	2	114
	Hordeum vulgare	26	
	Triticum urartu	4	
	Aegilops tauschii	4	
	Brachypodium distachyon	2	
	Zea mays	12	
	Sorghum bicolor	11	
	Panicum virgatum	4	
	Panicum hallii	3	
	Setaria italica	8	
	Oropetium thomaeum	2	
	Eragrostis tef	11	
Nelumbonaceae	*Nelumbo nucifera*	14	14
Pedaliaceae	*Sesamum indicum*	1	1
Solanaceae	*Nicotiana benthamiana*	8	8
Phrymaceae	*Mimulus guttatus*	2	2

*4 in Bur-0, 5 in C24, 6 in Can-0, 5 in Ct-1, 6 in E di-0, 6 in Kro-0, 4 in Ler-1

Table 5.2 | *PCP-B-like* and *UBL5* homologous genes retrieved from currently available completed and assembled genome databases for the phylogenetic analysis of these gene families.

Family	Species	*Number of PCP-B* homologues	Abbreviation	*Number of UBL5* homologues	Abbreviation
Brassicaceae	*Arabidopsis thaliana*	12	*AthB1-12*	*2*	*AthUBL5a-b*
	Arabidopsis lyrata	13	*AlyB1-13*	*2*	*AlyUBL5a-b*
	Boechera stricta	15	*BostrB1-15*	*2*	*BostrUBL5a-*
	Brassica oleracea	11	*BoB1-11*	*1*	*BoUBL5a*
	Brassica rapa	9	*BrapaB1-9*	*1*	*BrapaUBL5a*
	Capsella grandiflora	12	*CagraB1-12*	*1*	*CagraUBL5a*
	Capsella rubella	10	*CarubB1-10*	*1*	*CarubUBL5a*
	Eutrema salsugineum	7	*ThhalvB1-7*	*1*	*ThhalvUBL5a*
Malvaceae	*Gossypium raimondii*	5	*GoraiB1-5*	*0*	-
Poaceae	*Aegilops tauschii*	4	*AtauB1-4*	*2*	*AtauUBL5a-b*
	Panicum hallii	3	*PahalB1-3*	*2*	*PahalUBL5a-*
	Panicum virgatum	3	*PavirB1-3*	*5*	*PavirUBL5a-*
	Sorghum bicolor	14	*SobicB1-14*	*3*	*SobicUBL5a-*
	Triticum urartu	4	*TriurB1-4*	*0*	-
	Zea mays	12	*ZmB1-12*	*1*	*ZmUBL5a*

5.2.2 Phylogenetic analysis of PCPBL-encoding genes provides evidence for frequent gene duplication events and subfunctionalisation

Previous work reported in this thesis (Chapter 3, Wang L *et al*., 2016) revealed that PCP-B-like proteins are involved in early stages of the pollen-stigma interaction as well as early stages of seed development (Costa *et al.*, 2014). 15 PCPBL sequences were identified in the *Arabidopsis thaliana* genome. This observation suggested that the PCPBL-encoding gene family might have undergone frequent gene duplication events and significant functional diversification during its evolutionary history. To provide further evidence for this hypothesis, phylogenetic analysis of the 134 retrieved members of *PCPBL* gene family was performed. To exclude the possibility that some of the homologous genes were missed or multi-counted due to the quality of genome sequencing data sets, another gene family was selected as a reference in the selected genomes. Theses genes encode Ubiquitin-like protein 5 (UBL5), a family of non-reproductive proteins that are similar in size to the PCP-Bs (Table 5.2).

The reconciliation of the *PCPBL* gene tree with its species tree revealed an astonishing number of gene duplication events. There were 88 duplication events detected from the

reconciliation of the *PCPBL* phylogeny whereas only 14 were identified from the reconciliation of the *UBL5* phylogeny (Appendix 3, Figure S5.1). Interestingly, reconciliation of the *PCPBL* phylogeny showed 14 duplication events before the divergence of monocots and eudicots, revealing the occurrence of ancient gene duplication events in this gene family. In some species such as *Zea mays, Sorghum bicolor, Brassica oleracea and Boechera stricta*, the paralogous genes of *PCPBL* showed greater similarities to each other than the putative orthologous genes (Figure 5.1). There were 35 *PCP-B* homologous genes that fell into species-specific clusters, which indicated the occurrence of recent species-specific duplication events. In addition, the reconciled tree also revealed 41 and 27 lineage-specific duplication events in the Brassicaceae and Poaceae, respectively. Overall, these results provide evidence that the *PCPBL* gene family has been undergoing rapid evolution with widespread and frequent gene duplications throughout its evolutionary history, even within recently separated species.

The phylogeny of *PCPBL* genes can be interpreted as containing two main clades: a lineage-specific clade containing only members from the Poaceae and a clade principally containing members from the eudicots. The presence of Poaceae *PCPBL*s in the eudicot-dominated clade suggests that these genes arose before the divergence of monocots and eudicots. However, long-branch attraction, where large amount of change in the molecule could cause distantly related lineages appear to be closely related due to methodological artifacts, could potentially be an explanation for this this phenomenon (Bergsten, 2005). Importantly, the eudicot-dominated clade could be further divided into two clades: a 'PCP-B' clade and an 'ESF' clade that contained genes encoding PCP-Bs and ESFs (Embryo surrounding factors) (Costa *et al.*, 2014), respectively. This pattern of *PCPBL* phylogeny provided evidence for the subfunctionalisation or possibly even neofunctionalisation originating from gene duplication. The separation of sequences into clades containing PCP-Bs and ESFs also revealed putative orthologous genes in Brassicaceae, which made it possible to perform further molecular evolutionary analyses.

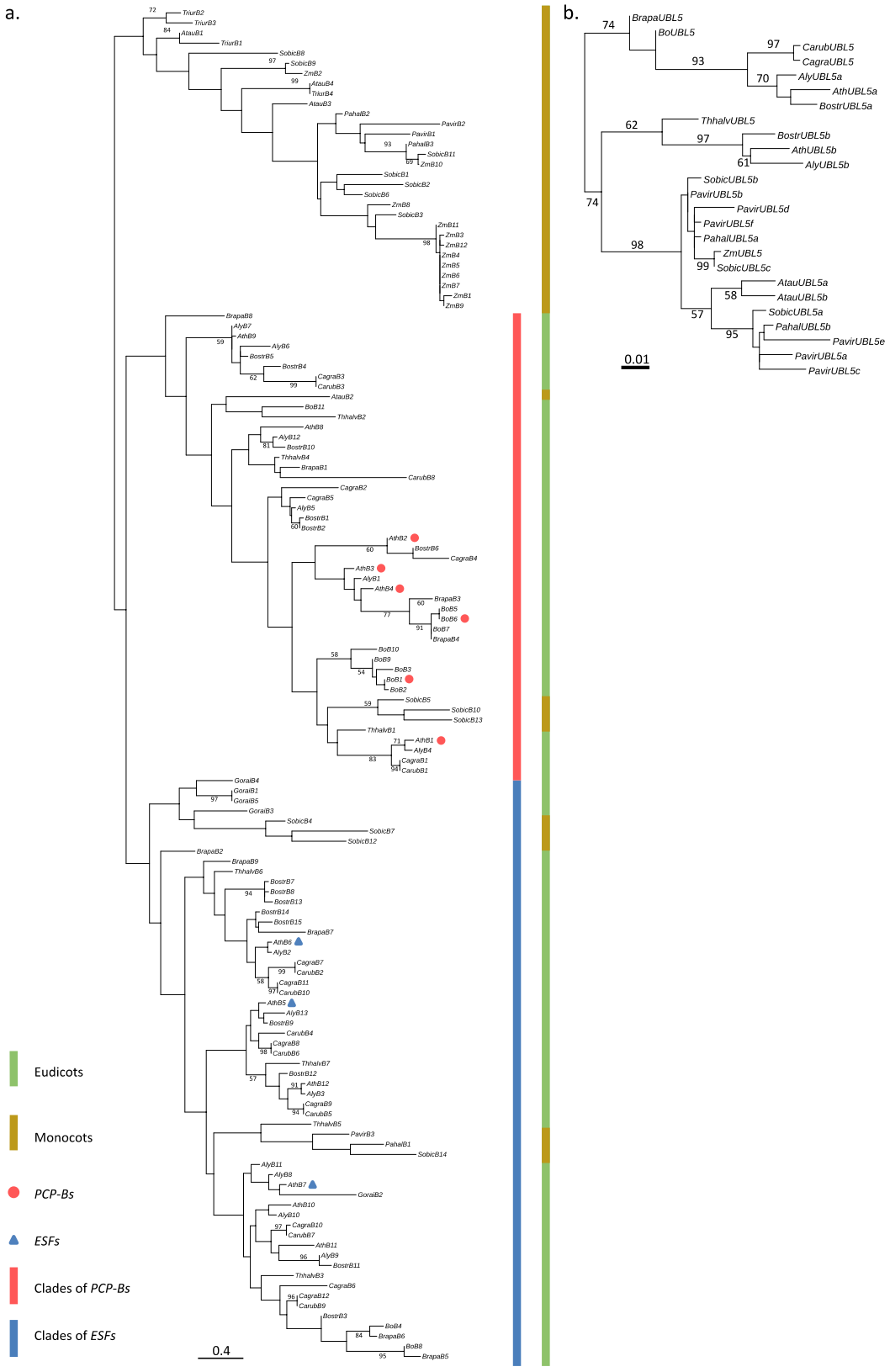
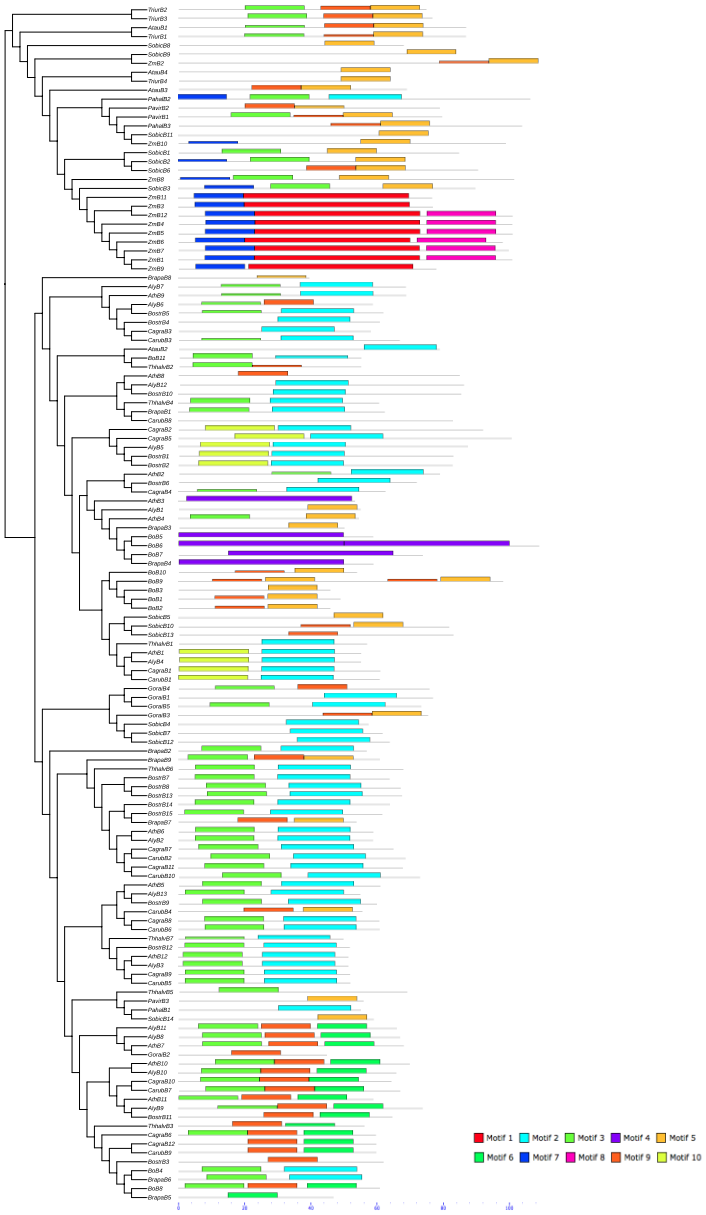


Figure 5.1 | Phylogenetic relationships of 134 *PCP-B* homologous genes from 16 species (a) and 25 UBL5 homologues from 14 species. This maximum likelihood (ML) tree was constructed by using the MEGA6 programme with the nucleotide sequences of PCPBL protein coding regions without signal peptides. The percentage bootstrap values (1000 re-samplings) higher than 50% are shown by interior branches. Branch length is scaled to the scale bar defined as 0.4 (a) and 0.01(b) nucleotide substitutions per codon. Gene abbreviations can be found in Table 5.1.

To investigate details of conservation and diversification of proteins encoded by *PCPBL* genes, the MEME programme was utilised to identify motifs in the mature protein sequences. Ten motifs were identified amongst 133 of the members of PCP-B-like protein family (Figure 5.3). Although four of these motifs (motifs 2, 3, 5 and 9) were identified based on their conserved cysteine residues, the similarities of their diversified regions lying between the cysteines were also considered. Motifs 1, 7 and 8 were identified as highly conserved regions due to the recent multiple duplications of *PCPBL* in *Zea mays*. Interestingly, motif 4 is unique in *At*PCP-Bγ and *Bo*PCP-B2, which strongly supports the similarity of their function in the early stages of pollination. The clustering of *PCPBL* genes with the same motifs provides further evidence for the reliability of the phylogenetic tree (Figure 5.1).



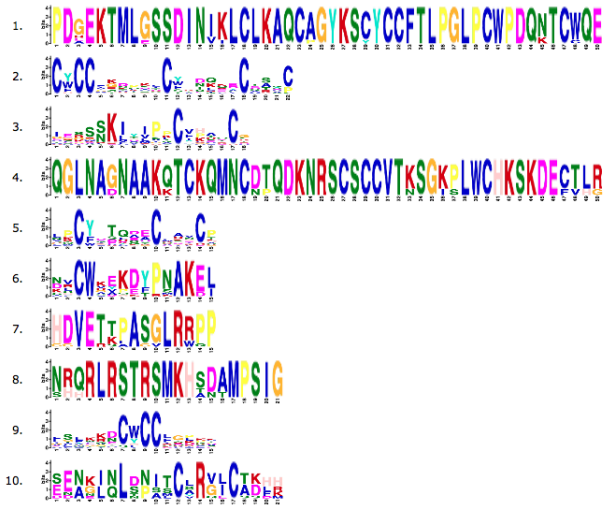


Figure 5.3 | Phylogenetic relationship of 134 *PCP-B* homologous genes and motifs identified from the amino acid sequences. The phylogeny is the same as for Figure 5.2 with branch length being ignored. Each of the sequences, except for 'CarubB8', has an *E*-value[a] less than 10. The block represents motif matches on the sequence where the position *p*-value[b] less than the significance threshold 0.0001(more significant). The height of a block indicates the significance of the match. For a motif block with a *p*-value greater than 1e-10, the block height is truncated and proportional to the negative logarithm of the position *p*-value. The scale numbers along the bottom of the figure represents amino acid positions.

a. The *E*-value of a sequence is the expected number of sequences in a random database of the same size that would match the motifs as well as the sequence does and is equal to the combined *p*-value of the sequence times the number of sequences in the database.

b. The position p-value is defined as the probability that a random sequence (with the same length and conforming to the background) would have a match to the motif under test with a score greater or equal to the largest found in the sequence under test.

5.2.3 Evolutionary patterns of *PCP-B*-like genes in *Arabidopsis thaliana*

The duplication mechanisms of *Arabidopsis thaliana PCPBL* genes can be partly revealed by their distribution on the Arabidopsis genome and their exon pattern. 15 *PCP-B* homologous genes (including three putative pseudogenes) with 11 of these forming two clusters located on Arabidopsis chromosomes 1 and 2 (Figure 5.4a), suggesting that they were the products of tandem duplications. Phylogenetically, *AthB3 (At2g16535/ AtPCPB-γ)* and *AthB4 (At2g16505/ AtPCPB-δ)* formed a well-supported terminal clade, which indicated that they were the result of a recent tandem duplication event (Figure 5.4b). However, the phylogenetic clade formed dominantly by the clustering genes on chromosome 1 also contained genes from other chromosomes, suggesting that these genes might be the consequence of more ancient tandem duplication events with the tandem array being subsequently disrupted and duplicated genes being distributed to different chromosomes during evolution. One of the putative pseudogenes, *AthB13 (At3g44212)*, lacked the characteristic intron of *PCP-B*s, suggesting that this gene is the result of retroposition. The ancestral genome of *Arabidopsis thaliana* has undergone three whole genome duplication (WGD) events after the divergence of angiosperms from seed plants, with the first round of WGD occurring before the divergence of eudicots and monocots whereas the most recent round occurred before the divergence of the *Brassica* and *Arabidopsis* genera (Jiao *et al.*, 2011). The reconciliation tree (Appendix 3 Figure S5.1) indicated about 50% of the duplication events that produced *PCPBL* genes happened after the divergence of Malvaceae and Brassicaceae. This suggested that the *PCPBL* genes not produced by tandem duplications might be the result of the third round of WGD prior to the emergence of *Arabidopsis*. However, none of the 15 *PCPBL* genes fell into the known non-hidden duplicated chromosomal blocks in the *Arabidopsis thaliana* genomes (Simillion *et al.*, 2002). This observation indicated either the collinearity of the syntenic blocks containing *PCPBL* genes was disrupted due to micro-rearrangements during their evolutionary history, or that multiple small-scale inter-chromosomal gene duplications have occurred.

To investigate whether any small-scale duplications have contributed to the expansion of the *PCPBL* gene family in *Arabidopsis thaliana*, gene syntenic comparisons were performed. Regions of the *Arabidopsis thaliana* genome that contained *PCPBL* genes were compared using GEvo (https://genomevolution.org/coge/GEvo.pl). Small fractions of

highly disrupted regions of collinearity neighbouring the *PCPBL* genes were revealed following the syntenic analysis (Appendix 3, Figure S5.3), which suggests a history of small-scale gene duplications (only one or a few genes are duplicated) or very ancient segmental duplications (Figure 5.4a) (Glover *et al.*, 2015).

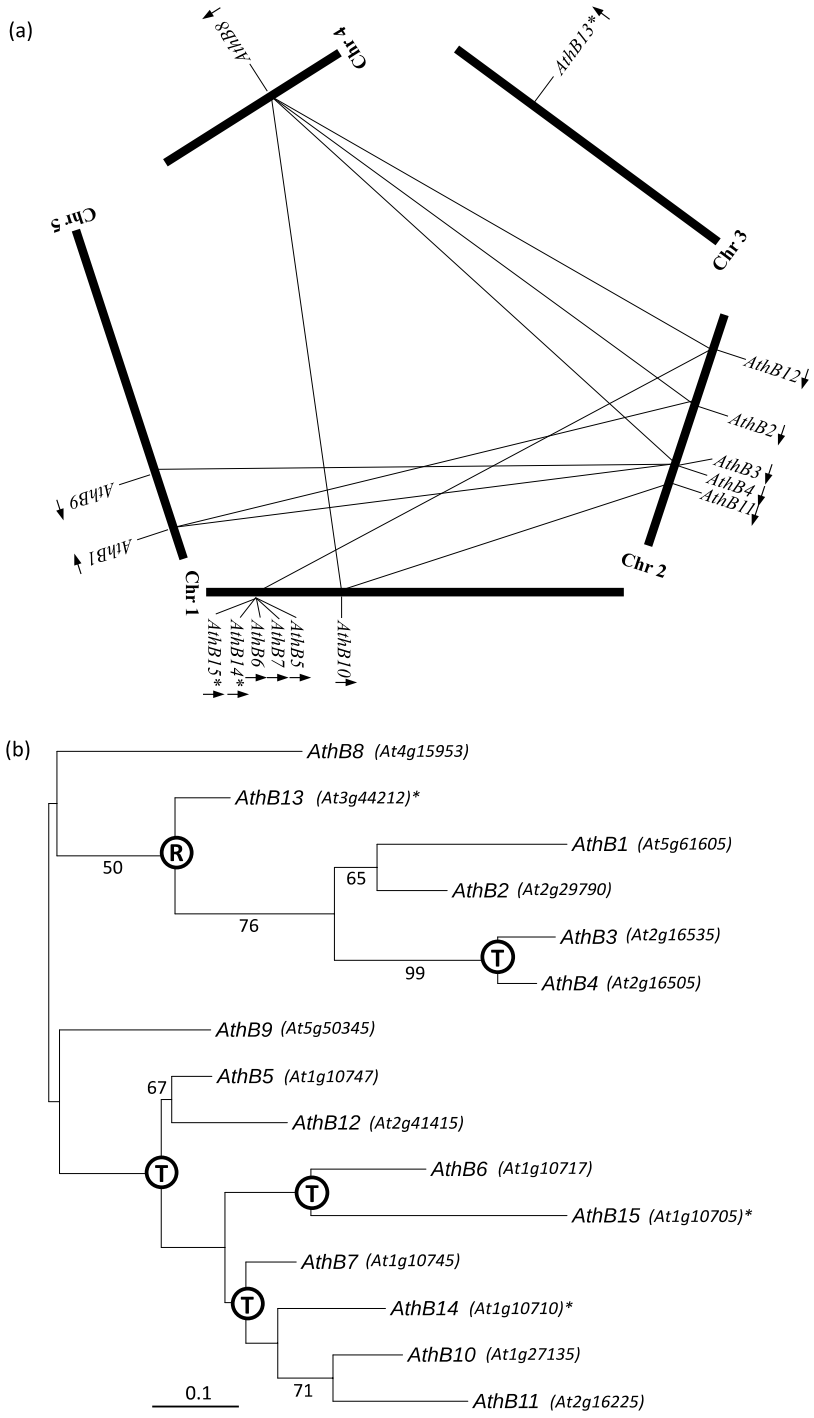


Figure 5.4 | Gene duplications and phylogenetic relationship of the 15 *PCP-B-like* genes (AthB) in *Arabidopsis thaliana*. (a) Distribution of 15 *PCP-B* homologues on *Arabidopsis thaliana* chromosomes and their putative syntenic relationships. The orientation of each gene is shown by an arrow. The putative pseudogenes are labelled with an asterisk *. The lines connecting *PCPBL* gene locations indicate the potential small-scale or segmental duplications. (b) Phylogenetic relationships of 15 *PCPBL* genes. Signal peptide encoding regions were excluded. The symbols T, and R on the tree nodes indicate when the tandem duplications and retroposition have occurred respectively. Branch length is scaled to the bar defined as 0.1 nucleotide substitutions per codon. The percentage bootstrap values (1000 re-samplings) higher than 50% are showed by interior branches.

To investigate the evolution of *PCPBLs*, molecular evolutionary analyses were carried out with paralogues in *Arabidopsis thaliana* and pairs of putative orthologues in *A. thaliana* and *A. lyrata*. To test for evidence of selection pressure acting along the branches of the *PCPBL* phylogenetic tree, a molecular evolutionary analysis was performed using the codeml programme in PAML package (Yang, 2007). The branch-specific free-ratio model allows an independent value of ω (ω = d_N/d_S ratio) to be generated for each evolutionary branch, which provides information on the variation of selection pressure on specific lineages. Evidence of positive selection was observed on 13 out of the 28 branches of the *PCPBL* phylogeny (Figure 5.5). Numbers of the nonsynonymous changes on the branches under positive selection revealed dramatically accelerated amino acid evolution.

The selection pressure of the *PCPBL* genes in *Arabidopsis thaliana* was also analysed by using putative orthologous genes from *Arabidopsis lyrata* (plants.ensembl.org). The d_N/d_S value of each pair of orthologues was calculated. The nucleotide substitution analysis along branches of Arabidopsis *PCPBL* phylogeny provided evidence for positive selection of some *PCPBL* gene copies. Interestingly, those branches with ω>1 (indicating positive selection) clustered in the *PCP-B* clade (branches b, c, d, e and f) and a portion (j, k, l and m) of the *ESF* clade (Figure 5.5), which contains genes known to be mostly involved in reproductive processes (Chapter 3, Wang L *et al.*, 2016; Costa *et al.*, 2014). When contrasting nonsynonymous substitution to synonymous substitution on the nucleotide sequences of putative orthologous *PCPBL* gene pairs in two *Arabidopsis* species, the pairwise ω= d_N/d_S did not show a value greater than one for each gene pair but did show large ω variation amongst gene pairs (Figure 5.5). However, despite the high d_N/d_S values on the branches showing positive selection, most of the ω values between orthologous gene pairs on these branches were lower than 1. These results indicate that purifying selection may be acting strongly on only a few sites along the *PCPBL* sequences during their evolution.

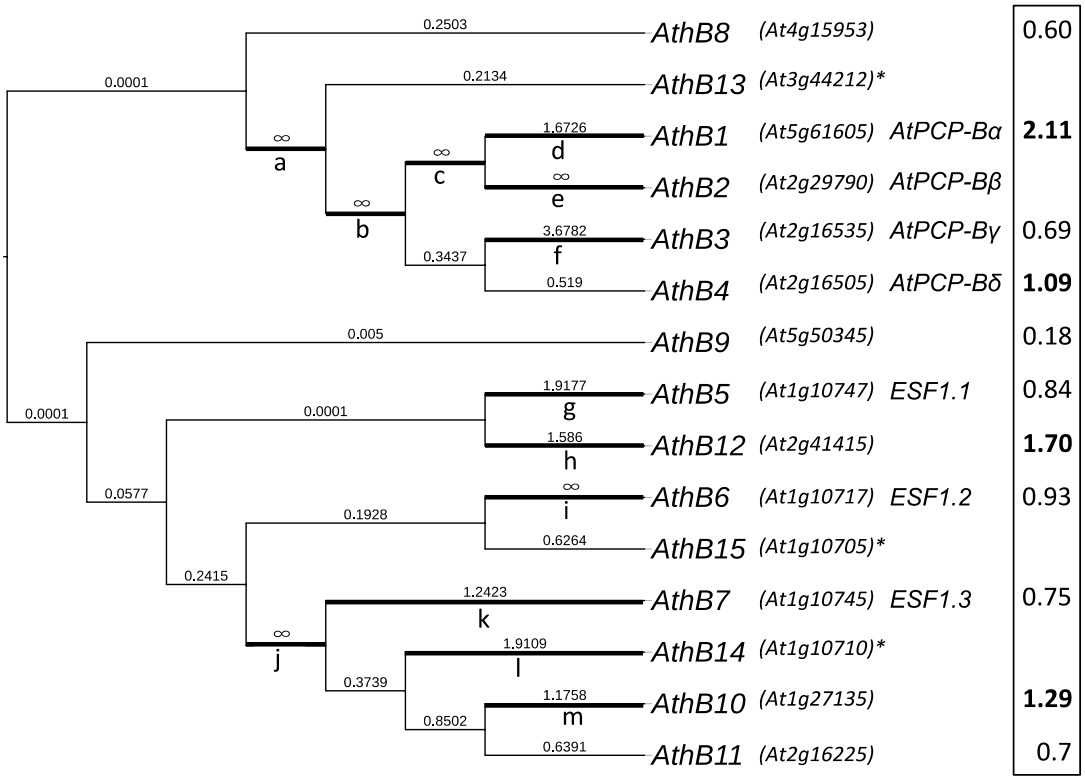


Figure 5.5 | Molecular evolution of PCP-B-like protein encoding genes in *Arabidopsis thaliana*. The branch-specific d_N/d_S values are shown adjacent to the branches, and were inferred by using the branch-specific test (free-ratio model) of codeml in the PAML package. The thick black lines indicate branches showing evidence of positive selection. The pairwise d_N/d_S values are shown by the leaf labels and highlighted by box, which were inferred with available putative orthologous genes in *Arabidopsis thaliana* and *Arabidopsis lyrata* by using YN00 in PAML. Values of pairwiseω greater than 1 are shown as bold. The phylogenetic tree presents the same topology as Figure 5.4 with branch length being ignored. The ω ratio and numbers of nonsynonymous and synonymous changes for each branch are as follows: a (∞; 81.3/0.0), b (∞; 91.5 /0.0), c (∞; 14.9 /0.0), d (1.6726; 132.8 /29.0), e (∞; 81.0 /0.0), f (3.6782; 36.2 /3.6), g (1.9177; 34.9 /6.6), h (1.586; 104.5 /24.1), i (∞; 76.6 /0.0), j (∞; 10.8 /0.0), k (1.2423; 35.4 /10.2), l (1.9109; 67.9 /13.0), m (1.1758; 50.8 /15.8).

5.2.4 Variable selective pressure among sites of putative AtPCP-Bγ orthologues and identification of amino acids under diversifying selection

To estimate the selection pressure on different sites of *AtPCP-Bγ*, of which single mutant line showed the most statistically robust defect during early pollen hydration, eight putative orthologous genes were selected from eight species in the Brassicaceae (Figure 5.6). The phylogeny and codon alignment of these eight orthologous genes was analysed using six different models within the codeml programme (Yang, 2007, Table 5.3). The comparison of a more complex model with a simple model by using a likelihood ratio test

(LRT) indicates which model fits the data better. Table 5.3 presents the LRT statistic $2\Delta \ln L$ for each pair of the model comparison. Model M0 assumes that all the sites are under the same selection pressure and performs the same ω ratio, while M3 allows for different ω ratios on each site. In M1 and M7, $0 < \omega_0 < 1$ is estimated from data and $\omega_1 = 1$ is fixed, while in M2 and M8, the sites are allowed to be positive selected (ω>1). There are four more parameters in M3 than in M0, which determines the degree of freedom (*df*) as 4. In this comparison, the LRT statistic $2\Delta \ln L = 46.87$ is much greater than the critical value (p= 0.01) $\chi^2_{1\%}$= 13.28 when $df = 4$. This result indicates that selective pressure varies across AtPCP-Bγ amino acid sites. The M2 added an additional ω that estimated to be 2.28. Although the log-likelihood didn't improve significantly, the *p*-value is very close to 0.05. Another comparison M7 and M8 showed significant improvement on the log-likelihood value (Table 5.3). In M3, the naive empirical Bayes (NEB) (Nielsen & Yang, 1998; Yang, 2000) was implemented, while in M2 and M8, the Bayes empirical Bayes (BEB) (Yang *et al.*, 2005) was available. In small data sets lacking information, the NEB approach could fail to account for sampling errors in maximum likelihood estimates of model parameters, including the proportions and ω ratios for the site groups, which may lead to unreliable posterior probability calculations. Though both methods calculate the posterior probabilities and identify positively selected sites, the BEB method is recommended due to its ability to account for sampling errors by applying a Bayesian prior probability distribution (Yang *et al.*, 2005). Thus, M8 was used as the final results for site selection analysis. The posterior probabilities of each site calculated under M8 were plotted in Figure 5.7. The diagram demonstrates the amino acid sites encoded *AtPCP-Bγ* with approximate mean of the posterior distribution ω>1, whereas the sites with posterior probability greater than 95% are highlighted. Among 40 sites inferred with posterior mean ω>1, twelve of them presented posterior probability of ω>1 higher than 95%. One site (37) presented posterior probability of ω>1 higher than 99%. These results provided evidence of different selection pressures for amino acid sites along the PCP-B sequences and strong positive selection on some of the sites. Amongst the 12 sites that showed no evidence of positive selection (posterior mean ω<1), eight of them are conserved cysteine residues that maintain the secondary structure of the molecule by forming disulphide bonds.

The sites inferred to be under diversifying selection (with their posterior probabilities based on M8) were mapped onto the predicted model of AtPCP-Bγ (see Figure 5.8 and

Chapter 3, Wang L *et al*., 2016). The 40 sites having a mean posterior ω>1 are scattered over the primary sequence, whereas the sites under strong positive selective pressure (posterior probability of ω>1 higher than 95%) are mostly clustered at the surface of the predicted secondary structure. Interestingly, when comparing the electrostatic potential surface model with the selection pressure model, a correlation between positive surface charge and diversifying selection was observed (Figure 5.8, Figure 9c). As the selection force acting on PCP-Bs could potentially be responses to a changing stigmatic protein, this result provides evidence that the sites under positive selection might be involved in the interaction between the pollen ligand and the stigmatic target molecules.

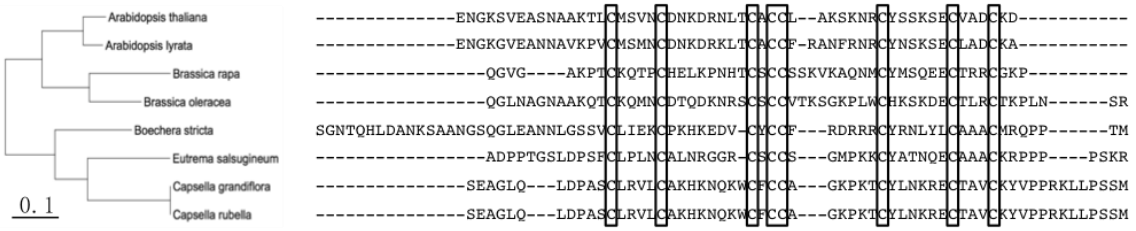


Figure 5.6 | The phylogeny and amino acid sequences alignment of eight putative *AtPCP-Bγ* orthologs. Branch length is scaled to the scale bar defined as 0.1 nucleotide substitutions per codon. Black boxes highlight the eight conserved cysteine residues.

Table 5.3 | Log-Likelihood values and Parameter estimates under models of variable ω ratios among sites

Model	p^a	In*L*	Parameters	$d_N/d_S{}^b$	Positive selected sites[c]
M0	1	-1281.794644	ω=0.972	0.972	None
M1	1	-1261.290963	p_0=0.13971, ω_0=0.00604 p_1=0.86029, ω_1=1	0.86	Not allowed
M2	3	-1258.361637	p_0=0.13234, ω_0=0 p_1=0.53675, ω_1=1 p_2=0.33091, ω_2=2.28	1.28	2, 3, 7, 9, 15, 18, 19, 20, 23, 26, 28, 29, 34, 35, 36, 37*, 38, 43, 46, 54
M3	5	-1258.358776	p_0=0.13203, ω_0=0 p_1=0.51994, ω_1=0.97403 p_2=0.34803, ω_2=2.22581	1.28	2, 3, 7, 9, 15, 18, 19, 20, 23, 26, 28, 34, 36, 37*, 38, 43, 46, 54
M7	2	-1261.536407	p=0.6520, q=0.00976	0.83	Not allowed
M8	4	-1258.360192	p_0=0.66133, p=0.03205, q=0.00976 p_1=0.33867, ω=2.25374	1.29	1, 2*, 3, 4, 5, 6, 7*, 9, 11, 12, 13, 14, 15*, 17, 18*, 19, 20, 22, 23, 24, 25, 26*, 27, 28*, 29, 31, 34*, 37**, 38*, 39, 40, 43*, 44, 45, 46*, 49, 50, 51, 53, 54*

a. Number of parameters in the ω distribution.
b. Average of ω over sites.
c. Sites under selection at the 99% level are labelled with **and at the 95% level are labelled with *.

Table 5.4 | Comparisons of data sets obtained by using multiple codeml models in PAML.

Comparison	2ΔIn*L*	*df*	χ^{2} [a]	*p*-value
M0 (one ratio) vs. M3 (discrete)	46.87	4	13.28	**<0.00001**
M1 (neutral) vs. M2 (selection)	5.86	2	5.99	0.053
M7 (beta) vs. M8 (beta & ω)	6.35	2	5.99	**0.042**

a. The critical value of $\chi 2$ distribution. For comparison of M0 and M3, the critical *P*-value is 0.01, while for comparison of M1/M2 and M7/M8, the critical *P*-value is 0.05.

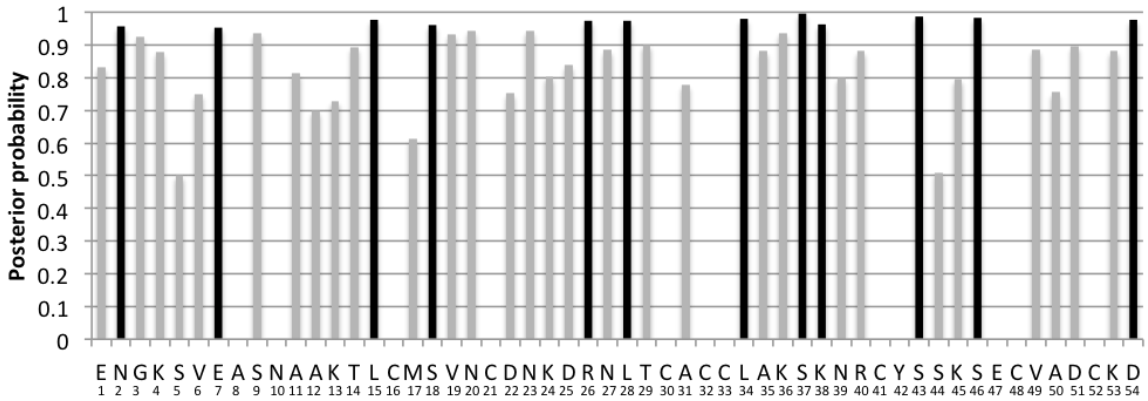


Figure 5.7 | Posterior probability for amino acid sites with posterior mean ω>1 based on model M8 in the codeml programme of the PAML package. The amino acid residues of AtPCP-Bγ are shown below the *x*-axis. The posterior probability greater than 0.95 were indicated by black bars.

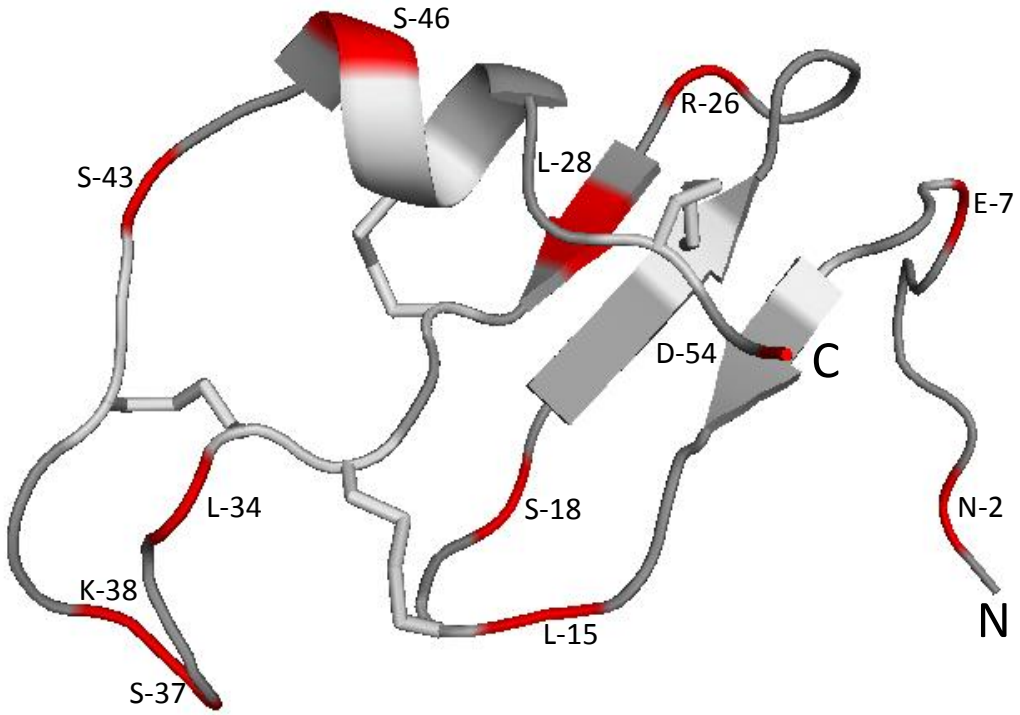


Figure 5.8 | Predicted AtPCP-Bγ structure with sites under diversifying selection. The amino acid sites under diversifying selection are labelled in red. The sites under purifying selection are labelled in light grey.

5.3 Discussion

Positive selection has been found to play important roles in a wide range of species. In humans, genes shown to be under positive selection have been discovered to be involved in diverse biological processes including host-pathogen interaction, reproduction, dietary adaptation, appearance, behaviour, brain development and sensory systems (reviewed in Vallender & Lahn, 2004). In *E.coli*, the primary target of positive selection was found to be the genes encoding cell surface protein due to their functions in the host immune/defence system by interacting with phages and bacteriocins (reviewed in Peterson *et al*., 2007). In plants, study of the genetic basis for mechanisms that contribute to reproductive isolation provides insights into speciation and angiosperm diversity.

Positive selection is a common feature of genes that underlie the formation of reproductive barriers (Swanson & Vacquier, 2002). As described in Chapter 3 (Wang L *et al*., 2016), a class of cysteine-rich proteins, the pollen coat protein B class (PCP-Bs) has been identified as important regulators of the early stages of pollination. The recently discovered and related Embryo Surrounding Factor 1 family (ESF1s) possess the same pattern of cysteine residues as PCP-Bs, highlighting the functional divergence of this protein family (Costa *et al.*, 2014). Both groups of proteins are involved in processes relating to reproduction in *Arabidopsis thaliana*. Thus, analyses of the phylogeny and molecular evolution of this protein family may shed new light on our understanding of interspecies reproductive barriers and even the origin of angiosperm diversity.

5.3.1 The collection of genes encoding PCP-B-like proteins

In this work, 280 PCP-B-like protein sequences were found to be present in six angiosperm families (based on all currently deposited plant genome data). The range of species with *PCPBL* genes present in their genomes revealed that this gene family might have originated before the divergence of monocots and eudicots. No gene sequences were found encoding PCP-B-like proteins in the basal angiosperms (Amborellales, Nymphaeales and Austrobaileyales, based on APG IV, Byng *et al.*, 2016) and more ancient taxa, including Bryophytes, seedless vascular plants and gymnosperms. In order to obtain the most reliable capture of PCP-B sequences from the available genome data, translated sequences were used as been found to be more effective than gene alignment for CRP predictions

(Silverstein *et al.*, 2007). However, as angiosperm genome sequencing, assembly and annotation is incomplete, it is likely that some of the *PCPBL* sequences have been missed. No *PCPBL* sequences were found to be presented in Charophytes (e.g. algae), yeast or animal genomes. Thus, providing a definitive time point for the origin of the *PCPBL* gene family remains unclear.

To investigate the phylogenetic relationship of the *PCPBL* gene family in angiosperms, the sequences from incomplete, unassembled genomes and pseudogenes were excluded from the analysis to prevent repetitive and poor alignment. The conservation of the signal peptide, cysteine residue positions and gene size provided confidence in the retrieved data set of 134 *PCP-B-like* (*PCPBL*) homologous genes. In addition to the previously built phylogeny of *PCP-B* homologous genes from two Arabidopsis species and two Brassica species (Chapter 3, Wang L *et al*., 2016), the phylogeny in this study included *PCPBL* genes from four more species in the Brassicaceae, one species in the Malvaceae and six species from the Poaceae.

5.3.2 The phylogeny of *PCPBL* genes indicates birth-and-death evolution

Although the incomplete availability of genome sequencing data puts limits on the study of the evolutionary history of the *PCPBL* gene family, a clear pattern of rapid birth-and-death evolution is still implied by the phylogeny built in this study. The reconciliation of the *PCPBL* gene phylogeny and the species tree revealed large numbers of gene duplication and loss events. Although definitive values for the rate of gene gain and loss could not be obtained due to the incomplete nature of genome sequence data, the comparison of reconciliation analyses between the *PCPBL* and *UBL5* gene families revealed great differences in the numbers of their gene duplication events.

The phylogeny built using 15 *PCPBL* genes in *Arabidopsis thaliana* provided a clearer picture of the duplication history of this gene family as it included three putative pseudogenes that were not included in the previous phylogenetic analysis presented in Chapter 3 (Wang L *et al*., 2016). The updated phylogenetic three, in addition to the distributions of *PCPBL* genes on the chromosomes (Figure 5.4a), revealed a larger cluster of tandemly duplicated genes around the *ESF1*s, which demonstrated that tandem duplication contributed to a large proportion of gene birth during the evolution of the

*PCPBL*s. Although the gene synteny analyses of genomic regions around *PCPBL*s did not reveal clear collinear blocks, it is possible that these regions have undergone subsequent rearrangements and inversions, presenting a disrupted syntenous relationship that masks their origins. Except for the putative pseudogene *At3g44212*, introns were present in all other *PCPBL* genes. Taken together these results suggest that apart from tandem duplications, ancient WGD and small-scale duplication events played an important role in the expansion of the *PCPBL* gene family in *Arabidopsis thaliana*, whereas only a very small contribution was from retroposition.

5.3.3 Evidence of positive selection on *PCP-Bs* and homologues revealed likely importance in reproductive isolation

Supported by empirical data (Chapter 3, Costa *et al.*, 2014; Wang L *et al*., 2016), the separation of PCP-Bs and ESF1s into different clades indicated that the *PCPBL* gene family has undergone functional diversification such as subfunctionalisation and neofunctionalisation (see section 1.1.4) during its evolutionary history. In addition, considering the high degree of sequence divergence observed amongst the PCPBL proteins together with the functional analysis of *PCP-B* genes in *Arabidopsis thaliana* (Chapter 3, Wang L *et al*., 2016), we hypothesised that the evolution of some *PCPBL* genes could be under positive selection.

There are three explanations that can be applied to the results of largely different level of purifying selection force amongst branches (Figure 5.5): 1. The gene functions are fixed and thus the sequences appear to be undergoing purifying selection; 2. The selection force on some of the duplicated gene copies is relaxed; 3. Positive selection is only applied to some codon sites of the genes. The patterns of positive selection amongst members of the *PCPBL* gene family may be explained by models proposed for the maintenance and evolution of duplicated gene copies (Innan & Kondrashov, 2010). For gene copies produced by ancient duplication events, their functional divergence may have become fixed, and following this they undergo a ‘preservation’ phase of evolution (Chapter 1, Figure 1.2) (Innan & Kondrashov, 2010). When we consider that *PCP-Bs* and *ESF1s* have evolved distinct functions, it is reasonable to envisage that the ancestors of these two classes of genes may have experienced neofunctionalisation, a concept that was originally

proposed by Ohno (Ohno, 1970). Based on his theory, when new copies undergo positive selection due to the functional redundancy, a new function might be acquired and maintained by selection. However, it was not clearly stated in this model how selection distinguishes between the duplicated and original copies. An extended version of Ohno's model, termed Duplication-Degeneration-Complementation (DDC), suggests that redundancy-induced relaxed selection operates on both copies reducing functional efficiency, which leads both copies to subfunctionalisation (Force *et al.*, 1999). Changes in the regulatory regions of the gene copies may contribute to differentiation of their expression patterns, and by example expression studies of *PCP-Bs* and *ESF1s* show that they are distinct (Chapter 3, Costa *et al.*, 2014). Another subfunctionalisation model, Escape from Adaptive Conflict (EAC), proposed that if the original gene preformed multiple functions, the duplication events provide the opportunity for each gene copy to become functionally specialised by positive selection (Hughes, 1994). All three models discussed above assumed that the duplication event does not affect fitness. However, a series of other models assume that the duplication brings benefits (Innan & Kondrashov, 2010). For example, the model of 'beneficial increase in dosage' proposed that if the fitness can be improved by the increase of gene dosage, the duplicated gene may be fixed through positive selection (reviewed in Kondrashov & Kondrashov, 2006). If the benefit from increased dosage after fixation is small, the selective pressure might be relaxed on the gene copies (Kondrashov *et al.*, 2002). In theory, this model may apply to three types of genes: genes that mediate organism-environment interaction, genes involved in protein-protein interaction and genes that are required in large dosage (Kondrashov *et al.*, 2002; Kondrashov & Koonin, 2004; Veitia, 2005).

The evidence of positive selection on recently duplicated *PCP-B* gene copies in Arabidopsis (*AtPCP-Bγ* and *AtPCP-Bδ*), may support the 'beneficial increase in dosage' model. Based on the phylogeny of *PCPBL* genes in *A. thaliana* and *A. lyrata* (Appendix 3, Figure S5.2), *AtPCP-Bγ* and *AtPCP-Bδ* clustered with the same *A. lyrata PCPBL* gene, suggesting that *AtPCP-Bγ* and *AtPCP-Bδ* are likely to be produced by a duplication event that occurred after the divergence of these two species. When contrasting the d_N/d_S of *AtPCP-Bγ* and *AtPCP-Bδ* with their common orthologous gene in *A. lyrata*, a difference in the ω values was observed: *AtPCP-Bδ* shows evidence of positive selection ($\omega = 1.09$) whereas *AtPCP-Bγ* shows evidence of purifying selection ($\omega = 0.69$) (Figure 5.5). Based on the fact that *A. thaliana* and *A. lyrata* diverged comparatively recently (10 million years

ago) (Beilstein *et al.*, 2010), it is possible that this duplicated gene pair is still in the fixation or fate-determination phase (Chapter 1, Figure 1.2) (Innan & Kondrashov, 2010). Thus, if this hypothesis is true, it corresponds to the 'beneficial increase in dosage' model and suggests that *AtPCP-Bδ* is the new copy. However, the d_N/d_S comparison analysis for eight putative orthologous *PCP-B* genes revealed significant variation in selective pressure amongst amino acid sites. It is noteworthy that the conserved cysteine and aromatic residues are indeed under purifying selection. These residues have been thought to be crucial for the maintenance of the protein function (Chapter 3, Costa *et al.*, 2014). The sites showing evidence of positive selection lying between the cysteines are thought to be important for the specificity of the pollen-stigma interaction mediated by PCP-Bs (Chapter 3, Wang L *et al.*, 2016). These sites under positive selection estimated by codeml M8 provided additional strong support for the hypothesis that the evolution of *PCP-Bs* is driven by selective pressure to establish reproductive barriers through their interaction with stigmatic targets. A very closely related example revealed a similar story: evidence of positive selection and the variation of selective pressure was discovered for the self-incompatibility determinants SCR and SRK (Guo *et al.*, 2011), as well as among the sites of Arabidopsis SCRL coding regions located between the conserved cysteine codons (Vanoosthuyse *et al.*, 2001). Structure-function studies have indicated that SCR can tolerate a degree of sequence variation, suggesting that the novel specificity between SCR-SRK interactions may have arisen from a gradually changing affinity between them (Chookajorn *et al.*, 2004; Guo *et al.*, 2011). Additionally gene conversion (a unidirectional transfer of DNA from one sequence to a highly homologous 'acceptor' that becomes identical) is an important factor for the evolution of duplicated genes, as it homogenises the variations among paralogous genes or repeating regions, leading to a reduction of divergence between duplicated copies (Teshima & Innan, 2004). Gene conversion can also allow for beneficial mutations to be shared between paralogues and to preserve gene function from deleterious mutations (Mano & Innan, 2008). Gene conversion was observed between *SRK* and *SLG* in *Brassica* (Sato *et al.*, 2002; Fujimoto *et al.*, 2006), *SRK* and its putative paralogous gene in *Arabidopsis lyrata* (Charlesworth *et al.*, 2003), as well as *SRK* paralogues and *ARK3* (*A. thaliana receptor kinase 3*) in *Arabidopsis thaliana* (Guo *et al.*, 2011). A classic example of this concerted evolution is seen in the animal reproductive protein VERL, and was described in Capter1 section 1.13. Similarly, we hypothesise that the driving force for positive selection of *PCP-B* genes might also derive from constantly changing stigmatic target(s). Similar to the evolution of SCR, PCP-Bs may also tolerate

sequence variation between the cysteine residues. First, interacting sites on a stigma target changes, resulting in a lower affinity with the PCP-B pollen ligands. If the gene conversion occurs among the paralogous gene copies of the stigmatic target, the mutation is tolerated and the pollen hydration regulated by this molecular dialogue still occurs. Especially in the self-compatible species *Arabidopsis thaliana*, once SI function has been lost, the altered affinity between ligand and target would have much less affect on fitness due to relaxed interspecies pollen competition. Thus, the selection force on the stigmatic target(s) is relaxed and concerted evolution occurs, which generates selective pressure on the PCP-B ligands that adapt to the constantly changing target(s).

This model also corresponds to the process of the establishment of reproductive barriers. After two populations have been isolated for a long time, the stigmatic targets evolved distinctively. The co-evolution of ligands and targets eventually establish the reproductive barrier. However, as pollen hydration does occur not only between closely related species but also between species in different genera, this very first checkpoint of pollination is likely to be a relatively weak reproductive barrier. Interestingly, AtLURE1 has been shown to be a female gametophyte attractant for growing pollen tubes and is involved in the erection of a reproductive barrier by virtue of its species specific interaction with its heteromer pollen receptor MDIS1-MIK (Wang T *et al.*, 2016). Heterologous expression of AtLURE1 in the synergid cells of *Torenia fournieri* permitted successful guidance of the growing of Arabidopsis pollen tube into the Torenia embryo sac, which confirmed the role of AtLURE1 in the formation of a reproductive barrier (Takeuchi & Higashiyama, 2012). However, the evolutionary analysis of orthologues in *A. thaliana* and *A. lyrata* found no evidence for positive selection on overall sequences or specific codon sites, while only neutral selection was estimated to be responsible for the rapid divergence of the paralogous genes that were duplicated after the divergence of these two species (Takeuchi & Higashiyama, 2012). These results suggest that positive selection is not necessarily the hallmark of the evolution of proteins involved in reproductive barriers due to the complexity of factors that may affect the selection force on each barrier. Nevertheless, evidence of positive selection can still be used for detecting potential molecular hotspots of signalling proteins in plant reproduction, the genetic basis of reproduction isolation and even speciation.

In conclusion, the evolutionary study of *PCP-Bs* and their homologues revealed a highly diverse gene family that has evolved to function as regulatory factors in sexual plant reproduction. The evidence of positive selection detected by statistical analyses in this study suggests that *PCPBL* gene function could potentially contribute to reproductive barriers. The variation of selective force across sites in *AtPCP-Bγ* and the correspondence between the positive surface charge and the codon sites under positive selection also provided a tantalising insight into the potential interacting hotspots of the pollen ligands and their stigmatic receptors.

Chapter 6 Proteomic analysis of pollen coat in *Arabidopsis thaliana* and *Brassica oleracea*

6.1 Introduction

As described in Chapter 1 (see section 1.4), amongst members of the Brassicaceae, pollen coat, which is deposited onto the outer exine wall during the late stages of pollen development, plays a central role in determining the compatibility of pollinations (reviewed in Edlund *et al.*, 2004). By being localised at the interface of the pollen-stigma interaction, coating-borne factors important for determining compatibility are rapidly released to engage in intercellular molecular communication. The main composition of pollen coat includes lipids, proteins, glycoconjugates and pigments (Piffanelli *et al.*, 1998). Ultrastructural studies have revealed physical and biochemical changes to the papilla cell wall elicited by contact with the pollen coat (Elleman *et al.*, 1992; Elleman & Dickinson, 1996), which suggests likely enzymatic action (Knox & Heslop-Harrison, 1970). The lipid-rich environment of pollen coat potentially stabilises regulatory molecules that act as 'keys' to 'unlock' resources from the stigma. Indeed it has been shown that defects in very long chain lipid synthesis in the pollen coat of *eceriferum* (*cer*) mutants leads to failure of pollen hydration (Preuss *et al.*, 1993; Hulskamp *et al.*, 1995; Fiebig *et al.*, 2000). Apart from the PCP-Bs, previous studies have shown that small cysteine-rich proteins (CRPs) carried in the pollen coat play important roles in the pollen-stigma interaction, including the male determinant SCR in *Brassica* SI system (Schopfer *et al.*, 1999; Takasaki *et al.*, 2000; Shiba *et al.*, 2001) and the PCP-As that were found to bind several stigmatic proteins (Doughty *et al.*, 1998; Takayama *et al.*, 2000).

The results presented in Chapter 3 (Wang L *et al*., 2016) of this thesis on the expression and function of Arabidopsis *PCP-B* genes provided further evidence for the importance of the pollen coat in the pollen-stigma interaction. A molecular evolutionary study of *PCP-B* homologous genes in the Brassicaceae also revealed that they have undergone rapid diversification and that positive selection on *PCP-B* codon sites is a feature of their evolutionary history (Chapter 5). There is now a large body of evidence demonstrating that the small proteins in pollen coat play centrally important roles in the pollen-stigma interaction by acting as factors that mediate cell-cell communication (Doughty *et al.*, 2000;

Takayama *et al.*, 2000). Thus characterising these small signalling molecules is now an urgent priority for illustrating the functions of pollen coat in the Brassicaceae and other species with 'dry' stigmas.

Previous proteomic profiling of pollen coat has been carried out in *Arabidopsis thaliana* (Mayfield *et al.*, 2001) and *Zea mays* (Wu *et al.*, 2015). In addition, a proteomic analysis of mature pollen grains from *Oryza sativa*, though not specifically directed at the pollen coat, likely contains information on pollen coat components (Dai *et al.*, 2006). An early study on pollen coat proteins in *Brassica napus* identified a group of oleosin-like proteins / glycine-rich proteins (GRPs) (Ross & Murphy, 1996; Murphy & Ross, 1998). The most recently published protein profiling of Arabidopsis pollen coat detected only ten proteins including five GRPs, two extracellular lipases (EXLs), two protein kinases and one potential EF-hand Ca^{2+} binding protein (Mayfield *et al.*, 2001). However, as discussed in Chapter 3 (Wang L *et al*., 2016), although pollen hydration defects have been discovered in *exl* and *grp* mutants (Mayfield & Preuss, 2000; Updegraff *et al.*, 2009), as for the *cer* mutants, these proteins are more likely affecting the biophysical properties of pollen coat rather than acting as signaling factors / ligands (Fiebig *et al.*, 2000). These previous proteomic analyses of pollen coat have failed to provide good coverage of small proteins and peptides with sizes <10 kDa, which may due to the scarcity of protein material and/or the low sensitivity and resolution of the detection systems employed. In recent years, proteomic profiling techniques have undergone rapid technological development delivering improvements in sensibility, cost effectiveness and data richness, which in turn has meant that these techniques have been utilised in many aspects of biological research. The high sensitivity of these detecting systems may thus enable us to overcome the problems encountered in previous proteomic analyses of pollen coat.

The removal of pollen coat from the outer exine surface of pollen grains had previously been described for *Brassica spp.* and provides a powerful technique for identifying proteins (free from contamination by protein from the pollen protoplast) that have functions in the earliest phases of the pollen-stigma interaction (Doughty *et al.*, 1993). In this project, the isolation of Arabidopsis pollen coat proteins has been a challenge due to low pollen yield and the high lipid content of the pollen coat mixture. Only about 75mg of mature pollen grains can be collected from one generation of Arabidopsis plants grown from 200 mg of seeds (approximately 6,000 plants). Here the aim was to develop this

technique for *A. thaliana* and to generate rich data sets for the pollen coat domain from both *Brassica oleracea* and *A. thaliana*.

In this chapter, liquid chromatography/mass spectrometry (LC-MS/MS) was utilised to analyse the protein components of pollen coat from *Arabidopsis thaliana* and *Brassica oleracea*. Proteome profiling of the pollen coat from these two species revealed a strikingly large number of small cysteine-rich proteins that previously have not been reported as pollen coat components. The richness of the data sets demonstrate the great sensitivity of this approach, and provides an excellent source of targets that can be screened for their potential roles in the pollen-stigma interaction and the basal compatibility system.

6.2 Results

6.2.1 The isolation of protein components from pollen coat and the separation of proteins by HPLC

The pollen coat of Arabidopsis was extracted using cyclohexane and the mixture was left to dry on a glass slide. As had been described from previous trials in the lab (J. Doughty unpublished observations), the resulting fraction was observed to produce a yellow oily crusted mass rather than the honey-like semifluid for *Brassica oleracea.* This was problematic and complicated sample collection and led to sample loss. Thus, an aim of part of this project was to develop a modified technique that could maximise the protein yield. This involved the removal of cyclohexane in a freeze-dryer (see section 2.10.2).

The freeze-dried pollen coat extracts could not readily be dissolved in water and sonication was carried out to homogenise the mix which yielded a milky homogeneous suspension with foam on the liquid surface. Intense centrifugation (21,000***g***, 6 x10 min) was required to separate the lipidic phase from the pollen coat protein-containing aqueous phase and the latter was carefully collected after each centrifugation step and recentrifuged until the aqueous phase was a colourless clear solution.

Pollen coat protein extracts were analysed by SDS-PAGE followed by silver staining (Figure 6.1a). This confirmed that pollen coat proteins had successfully been isolated from both *Brassica oleracea* and *A. thaliana* and also revealed that their respective protein profiles were quite different from one another, at least with respect to the relative band intensities. For example scrutinising bands between 20 kDa to 25 kDa, for *B. oleracea*, a very abundant band at 20 kDa can be observed whereas the most abundant band of Arabidopsis pollen coat shows at around 24 kDa. When compared with the previously published Arabidopsis pollen coat proteome (Figure 6.1b, Mayfield *et al*., 2001), our SDS-PAGE analysis (Figure 6.1a) revealed more abundant proteins with small sizes (<20 kDa) but less abundant bands at larger sizes (>40 kDa). Comparison of these two gel profiles did reveal many commonalities between the two samples with respect to identifying some of the most abundant Arabidopsis pollen coat proteins such as GRP16 (22 kDa), GRP18 (21.5 kDa), GRP14 (18.5 kDa), GRP19 (18 kDa) and EXL4 (37.9 kDa) (Figure 6.1, 6.2), and these were further confirmed by the proteomic analysis of Arabidopsis pollen coat (Appendix 4 Table S6.1). Following SDS-PAGE analysis, Reverse Phase High Performance Liquid Chromatography (RP-HPLC) was carried out with the aim of developing a technique to readily isolate the protein components for subsequent protein identification, future bioassays or binding assays (depending on the future direction of the project). A preliminary experiment was carried out with pollen coat components from *Brassica oleracea* to estimate the resolution of this approach. The results demonstrated, as expected, a large number of protein components evidenced by the complex peak pattern obtained following HPLC (Figure 6.2a). Peaks were manually collected and analysed on SDS-PAGE (Figure 6.2b). The majority of larger proteins (those > ~15 kDa) were poorly recovered by RP-HPLC as can be ascertained by comparing the protein profile of the crude pollen coat extract (Fig 6.1) and peak analysis post-HPLC (Fig 6.2b). However the majority of small proteins (< ~15 kDa) were recovered with high efficiency following RP-HPLC though these were poorly separated by the 4-20% SDS-PAGE gel used for this broad analysis (Fig 6.2b). Even though the entire collected product from each peak was loaded on the gel, only the most abundant of the larger proteins could be detected. In addition, although this approach for separation of pollen coat proteins could be useful for some applications it was considered to lack the efficiency and sensitivity required for a thorough proteomic survey of this pollen domain. An alternative approach was taken and is described in the following section (6.2.2). In summary the isolation of pollen coat proteins

was successfully achieved for both *A. thaliana* and *B. oleracea* and their respective protein profiles were revealed to be strikingly different as determined by SDS-PAGE.

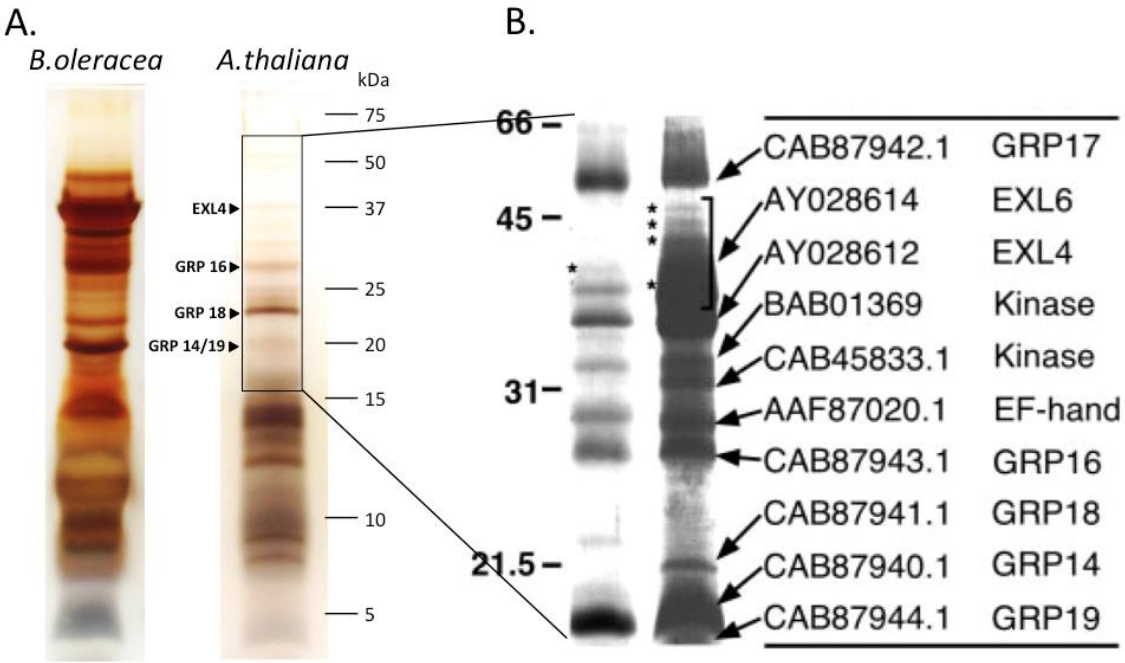


Figure 6.1 | Pollen coat protein extracts from *Arabidopsis thaliana* and *Brassica oleracea* (A) compared with previously published Arabidopsis pollen coat proteome (B) (Mayfield *et al*., 2001). 5% of total amount of extraction from each sample (Arabidopsis, 150 mg of pollen grains; *Brassica oleracea*, 75 mg of pollen grains) was loaded on SDS-PAGE followed by silver staining. Comparison of these two gel profiles is indicated by a black box. The possible corresponding protein bands are indicated by black triangles.

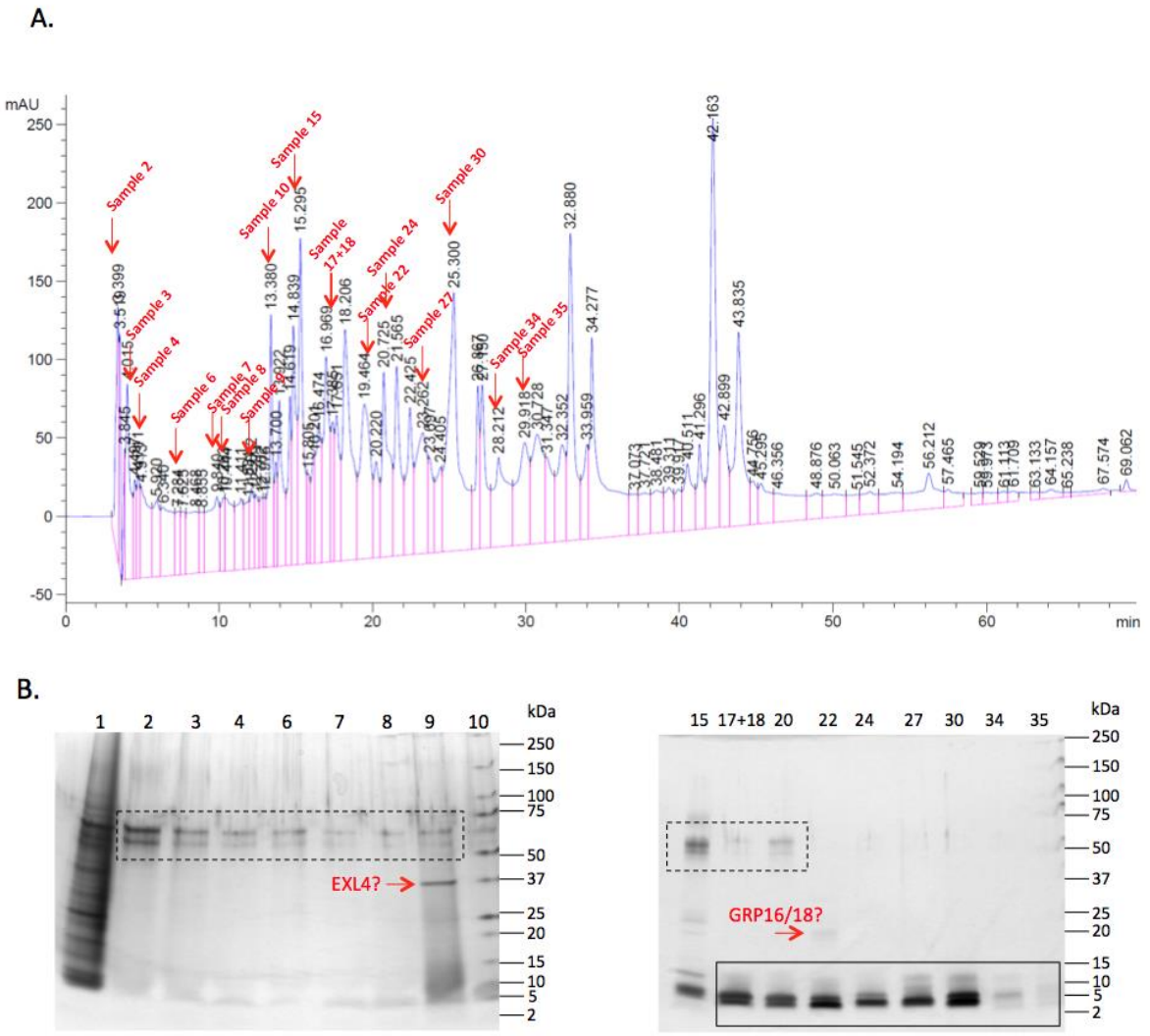


Figure 6.2 | HPLC analysis of protein extraction from *Brassica oleracea* pollen coat. A. Detected peaks with labelled sample names (wavelength 215 nm). B. Analysed peaks on SDS-PAGE gels. Mini PROTEAN® TGX Precast gel (4-20%), silver stained by SilverXpress® Kit. Samples: (HPLC 2014_04_24) collected peaks1-4, 6-10, 15, 17+18, 20, 22, 24, 27, 30, 34, 35. Proteins identity estimated by their sizes based on the corresponding Arabidopsis pollen coat protein data set are indicated by red arrows. EXL4: Extracellular lipase 4 (37.63 kDa), GRP16: Glycine-rich protein 16 (22.30 kDa), GRP18: Glycine-rich protein 18 (21.48 kDa). The solid lined box indicates bands of small pollen coat proteins recovered with high efficiency by HPLC. The dashed line boxes indicate possible keratin contamination.

6.2.2 Gene families and classes of CRPs from the Arabidopsis and Brassica pollen coat proteome

The previous approach involving HPLC separation of the pollen coat protein complement showed poor sensitivity and yielded an incomplete data set due to the scarcity of protein material, overlapping retention time of molecules and limitations of reverse-phase HPLC separation of larger proteins. To achieve the highest possible resolution and sensitivity, LC-MS/MS was performed to permit detection of low abundance protein components of the pollen coat and produce more complete data sets. This approach was selected due to several advantages: 1. Crude/complex samples could be loaded directly onto the system without gel separation thus preventing sample loss. 2. The detection system was highly sensitive and thus molecules with very low abundance could be detected. 3. Proteins having similar sizes were pre-separated by HPLC as an integral component of the mass spectrometry system thus providing higher resolution than the previous approach used by Mayfield *et al*. (2001).

Proteomic analyses of pollen coat proteins from both *Arabidopsis thaliana* and *Brassica oleracea* (Appendix 4, Table S6.1 and S6.2) revealed profiles with a size range of 5-254 kDa with a large proportion of the samples being made up of small proteins. For Arabidopsis, 271 proteins were obtained from merged data sets (Appendix 4, Table S6.1) of two LC-MS/MS analyses of the pollen coat, of which 42% had a molecular weight below 20 kDa. For *Brassica oleracea*, 227 proteins (Appendix 4, Table S6.2) were identified as pollen coat protein components, of which 48% had a molecular weight below 20 kDa (Figure 6.3). Among the proteins with a molecular weight below 15 kDa, most of them were found to be cysteine-rich. Only 3.8% of detected proteins had a molecular weight above 100 kDa. The distributions of protein sizes from Arabidopsis and *Brassica oleracea* pollen coat were similar, which indicated that the sensitivity of this proteomic analysis approach was constant.

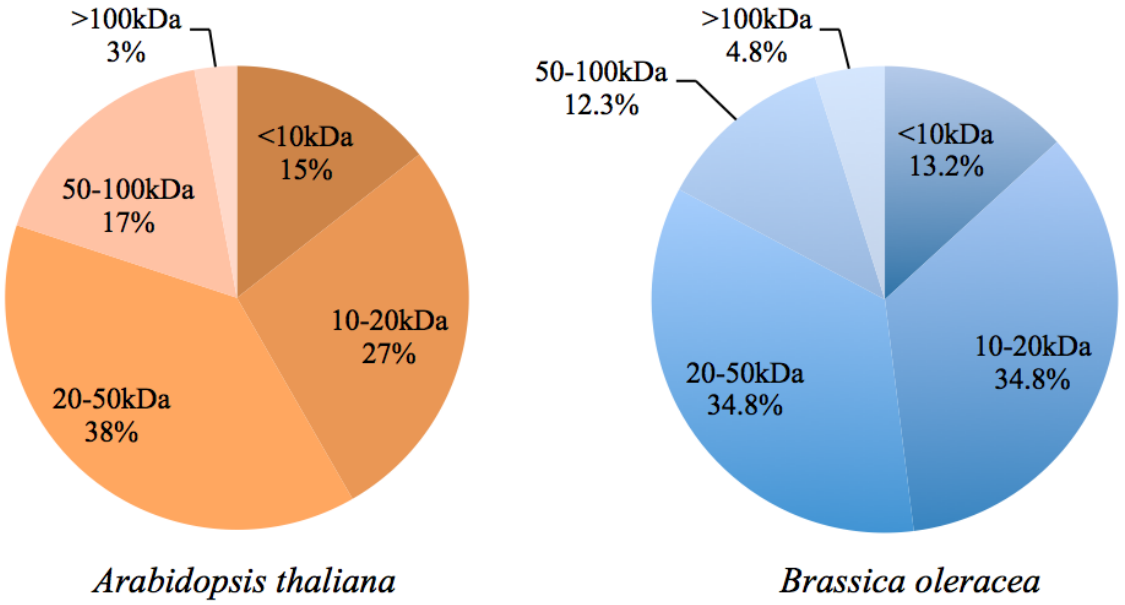
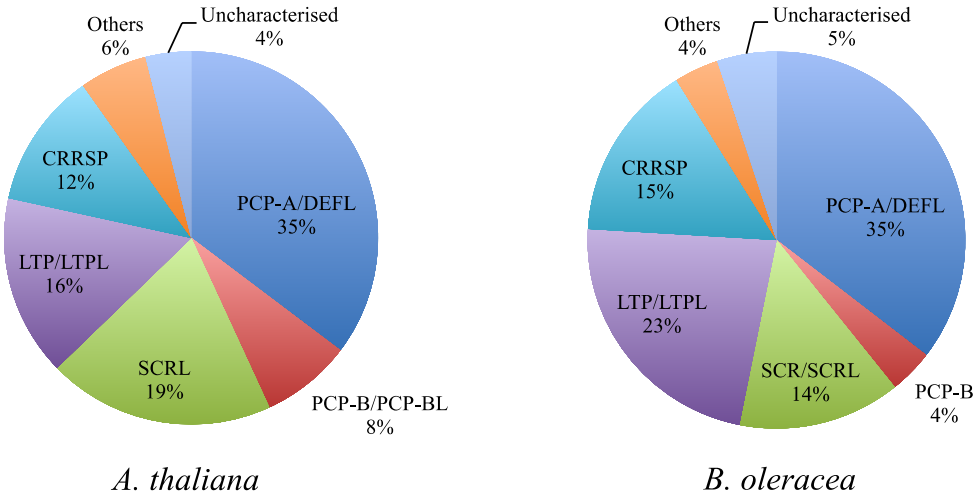


Figure 6.3 | Distributions of protein sizes in pollen coat proteomics of *Arabidopsis thaliana* (merge of two data sets) and *Brassica oleracea* (one data set) gained from LC-MS/MS.

A previous proteomic analysis of the Arabidopsis pollen coat (Mayfield *et al.*, 2001) revealed some of the most abundant protein components, including one EF-hand protein, two extracellular lipases (EXL4 and EXL6), five oleosin/ GRP proteins (GRP14, GRP16, GRP17, GRP18 and GRP19) and two kinases. The gene structure analysis revealed clusters of gene loci encoding six EXLs (EXL1-6) and six GRPs (GRP14, GRP16, GRP17, GRP18, GRP19 and GRP20) (Mayfield *et al.*, 2001). In this study, we detected the presence of EXL5 and GRP20 in the Arabidopsis pollen coat, proteins that were not detected in the research of Mayfield *et al.* (2001) – this highlighted the high sensitivity of our approach.

As mentioned in Chapter 1 (section 1.3.3), cysteine-rich proteins have been identified as important regulators of sexual reproduction amongst various angiosperm species. Our proteomic analyses of Arabidopsis and *Brassica oleracea* pollen coat revealed 130 cysteine-rich proteins were present in the pollen coat extracts, of which 51 were from Arabidopsis and 79 from *Brassica oleracea*. Based on their cysteine residue patterns, these pollen coat CRPs can be divided into classes, of which the five largest classes shared between the two species are PCP-As/DEFLs, PCP-Bs, SCRs, non-specific lipid-transfer proteins (nsLTPs) and cysteine-rich repeat secretory proteins (CRRSPs) (Table 6.1, Figure 6.4). The relative proportions of CRP classes in the two pollen coat proteomes were similar. PCP-A-like proteins share a similar cysteine residue pattern with DEFL and LCR proteins,

and these made up the largest proportion of CRPs. Similar proportions of SCR-like proteins (SCRL), nsLTP-like (nsLTPL) and CRRSPs were present in both pollen coat proteomes (Figure 6.4). Importantly, PCP-Bs were detected in both the Arabidopsis and *Brassica oleracea* pollen coat, which provided further support for the hypothesis that PCP-Bs function in plant reproductive signalling (Chapter 3, Wang L *et al*., 2016).



A. thaliana *B. oleracea*

Figure 6.4 | Classifications of CRPs detected following LC-MS/MS proteomic analysis of *Arabidopsis thaliana* and *Brassica oleracea* pollen coat. PCP, pollen coat protein; SCR(L), S-locus cysteine-rich (like) protein; LTP(L), lipid transfer protein (like); DEFL, defensin-like protein; CRRSP, cysteine-rich repeat secretory protein.

Table 6.1 | Classification of CRPs present in the proteomes of Arabidopsis and *Brassica oleracea* pollen coat. PCP, pollen coat protein; DEFL, defensin-like protein; SCR(L), S-locus cysteine-rich (like) protein; nsLTP, non-specific lipid transfer protein; CRRSP, cysteine-rich repeat secretory protein; GASA, giberellic acid stimulated in *Arabidopsis*; ECA 1, early culture abundant 1; RALF, rapid alkalinisation factor.

CRP classes[a]	***A.thaliana***[b]	***B.oleracea***[c]	**Sizes**[d]	**Cysteine-patterns**[e]
PCP-A/DEFL/LCR	18	28	54-101	**C**X(3-12)**C**X(4-8)**C**XXX**C**X(5-15)**C**X(3-16)**C**X**C**X(1-4)**C**
PCP-B	4	3	76-82	**C**XXXX**C**X(6-9)**C**X**C****C**X(6-9)**C**X(6-12)**C**XXX**C**
SCR/SCRL	10	11	73-108	**C**X(9-10)**C**X(7-8)**C**X(13-23)**C**X(1-2)**C**X(12-27)**C**X**C**(3-7)**C**
nsLTP	8	18	91-265	**C**X(6-9)**C**X(10-16)**C****C**X(8-19)**C**X**C**X(12-25)**C**X(5-14)**C**
CRRPS	6	12	253-264	**C**X(45-56)**C**X(8)**C**XX**C**X(11)**C**X(13)**C**X(75-84)**C**X(6-8)**C**XX**C**X(22-25)**C**
Gibberellin-regulated protein GASA	2	1	89-94	**C**XX**C**XXX**C**X(8)**C**XXX**C**XX**C****C**XX**C**X(11)**C**X**C**X(12)**C**
ECA1-like gametogenesis related family protein	1	1	97-104	**C**X(10-13)**C**X(18)**C****C**X(8-9)**C**X(18)**C**
RALF-like	0	1	73	**C**X(7)**C**X(11)**C**X(5)**C**
Uncharacterised	2	4	125-265	-

a, The CRP class names were defined by the UniProtKB annotated members in each class.
b, c, The numbers of proteins in each class based on their cysteine residue patterns.
d, The range of amino acid residue numbers (with signal peptides) in each class.
e, Conserved cysteine residues that define each CRP class. The bold C represents a conserved cysteine. X represents any amino acid; numbers in brackets represents the range of variable residue numbers between conserved cysteines.

6.2.3 Distributions of genes encoding pollen coat CRPs on genomes of *A. thaliana* and *B. oleracea*

The distributions of genes encoding the five largest pollen coat CRP classes on chromosomes of *Arabidopsis thaliana* and *Brassica oleracea* revealed multiple gene clusters that were likely produced by tandem gene duplications. The largest gene cluster encoding pollen coat CRPs in the Arabidopsis genome consists of five CRRSPs on chromosome 3 (Figure 6.6). Other pollen coat CRP-encoding genes were also observed to be likely tandem duplications, including two pairs of genes encoding PCP-A-like proteins (PCPALs) on chromosome 2, a pair of genes encoding two SCR-like proteins (SCRLs) on chromosome 4 and a pair of genes encoding two of the previously characterised PCP-Bs (PCP-Bγ and PCP-Bδ, Chapter 3, Wang L *et al*., 2016) on chromosome 2. For *Brassica oleracea*, multiple tandemly duplicated gene clusters were also observed (Figure 6.7). The presence of gene clusters encoding CRRSPs, PCPALs and SCRLs on both genomes suggests that these gene families have an origin in a common ancestor of *Arabidopsis* and *Brassica*. For example, comparison of genome regions encoding CRRSPs clusters in *A. thaliana* (chromosome 3) and *B. oleracea* (chromosome 5) revealed highly syntenic blocks (Figure 6.5), demonstrating duplication histories that generated CRRSP-coding gene clusters before the divergence of *Arabidopsis* and *Brassica.*

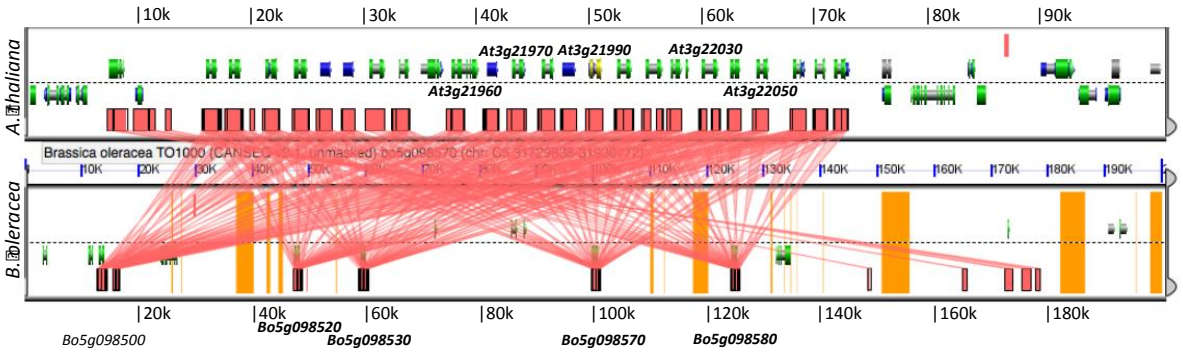


Figure 6.5| Comparison of genome regions encoding CRRPSs in *A. thaliana* and *B. oleracea.* Red wedges show connection of syntenic regions shown as red blocks. The genes encoding CRRSPs detected by LC-MS/MS were labelled by gene annotation in bold.

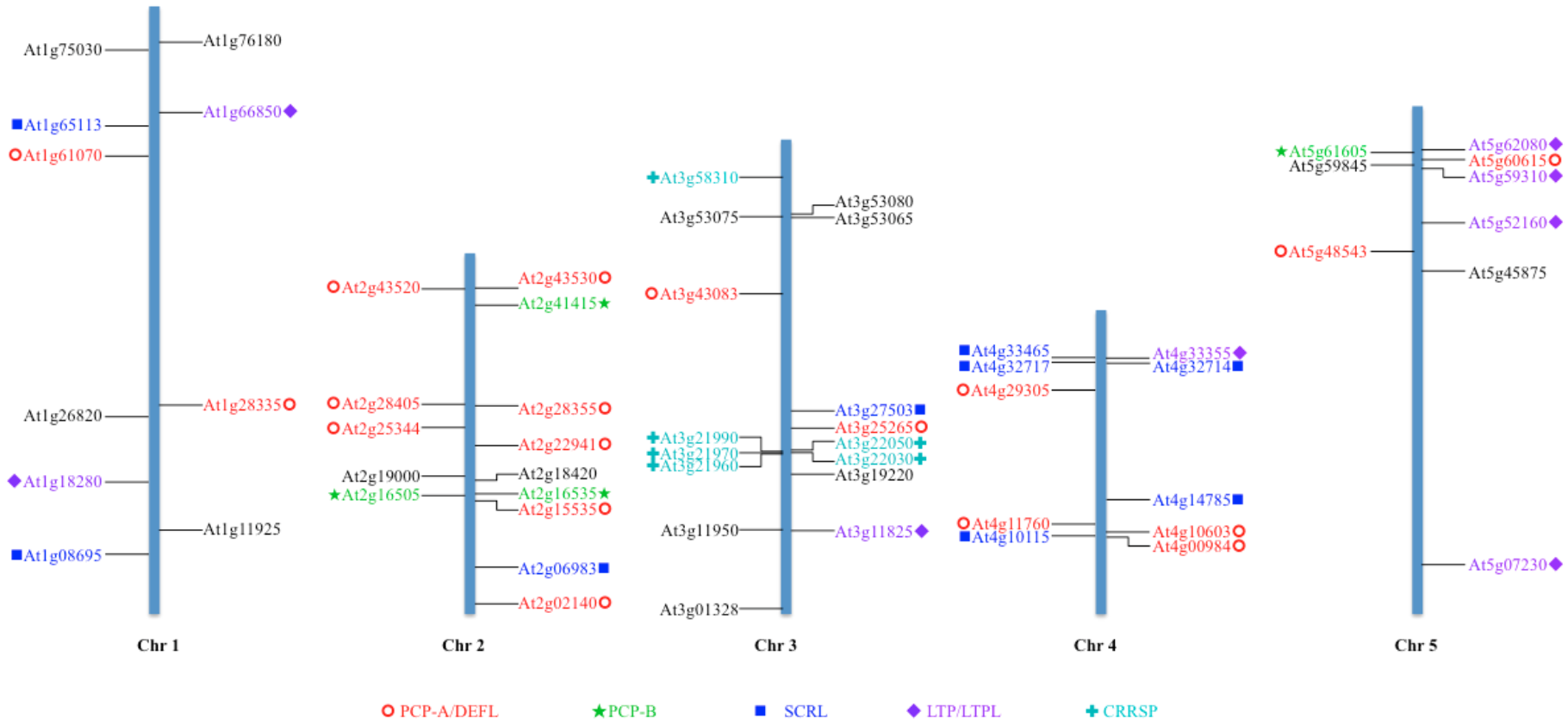


Figure 6.6 | Distributions of pollen coat CRP classes on chromosomes of *Arabidopsis thaliana*. The five largest classes are colour coded.

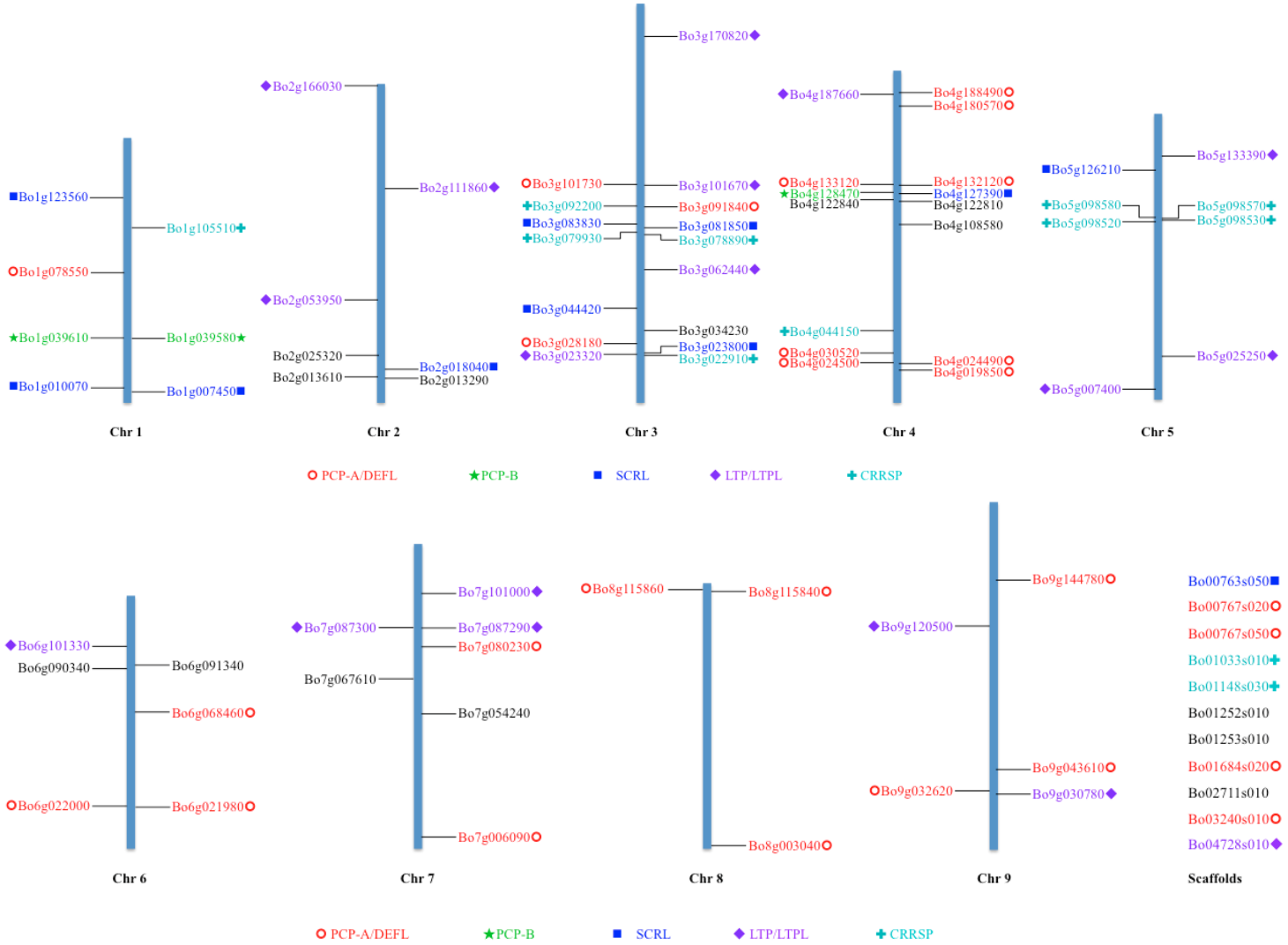


Figure 6.7 | Distributions of pollen coat CRP classes on chromosomes of *Brassica oleracea*. The five largest classes are colour coded.

6.2.4 Gene ontology of pollen coat proteome in *Arabidopsis thaliana*

To obtain an overview of the possible roles played by pollen coat protein components, Gene Ontology Consortium data was utilised (Ashburner *et al.*, 2000; Blake *et al.*, 2015). Due to the availability of gene annotation in the GO database, this analysis was only performed with the Arabidopsis pollen coat proteome. Gene functions were classified in this analysis in two ways: 1. Molecular function, describing activities that occur at the molecular level, which can be performed by individual gene products or complexes formed by gene products. 2. Biological process, describing a series of events mediated by one or multiple assembled molecular functions. The gene ontology (GO) enrichment analysis assigned putative function to 61% of pollen coat proteome in Arabidopsis leaving a remarkably large number of proteins with unassigned functions (Figure 6.8, 6.9). The 'molecular function' as defined by gene ontology PANTHER classification system (PANTHER GO-slim, see Chapter 2, 2.6.6) revealed two dominant functional classes, these being 'catalytic activity' and 'binding'. 46 proteins were identified that fell into the 'catalytic activity' class (Table 6.2), whereas 23 proteins were identified as binding proteins (Table 6.3, Figure 6.8). Genes classified by 'biological process' revealed that a large proportion of proteins were likely involved in metabolic (40 proteins) (Table 6.4) and cellular processes (34 proteins) (Table 6.5, Figure 6.9). However, many of the proteins fell into categories annotated as being involved in intracellular activities, such as GTP-binding proteins (Q3E6Q3, O04834 in Table 6.2), profilin (Q42449, an actin-binding protein) and glycosyltransferase (F4IBV4), indicating possible contamination of the pollen coat fraction by cytoplasmic proteins. However another likely explanation for their presence in pollen coat is that they are simply relics of tapetal dissolution and end up transferred to the pollen coat having no actual function in this pollen domain (Quilichini *et al.*, 2014). By considering that the pollen coat is at the interface between pollen and papilla cells, the detection of some proteins supports its importance in pollen-stigma interaction. For example, esterase has been reported to be crucial for the breakdown of the stigmatic cuticle to allow pollen tube penetration (Evans *et al.*, 1992). Cysteine protease inhibitors were also identified and these may act as 'protectors' preventing other pollen coat proteins from degradation, although they may also exhibit antimicrobial activity (Kim *et al.*, 2009). Interestingly, none of the CRPs detected in the Arabidopsis pollen coat were annotated by the PANTHER classification system (included in unmapped ID, Figure 6.8, 6.9), which reveals the lack of empirical evidence for the functions of these gene products and their

potential importance as signalling molecules in pollen-stigma communication. Thus, further studies need to focus on these CRPs and other uncharacterised pollen coat proteins. This should shed new light on the pollen surface regulators of reproductive signalling.

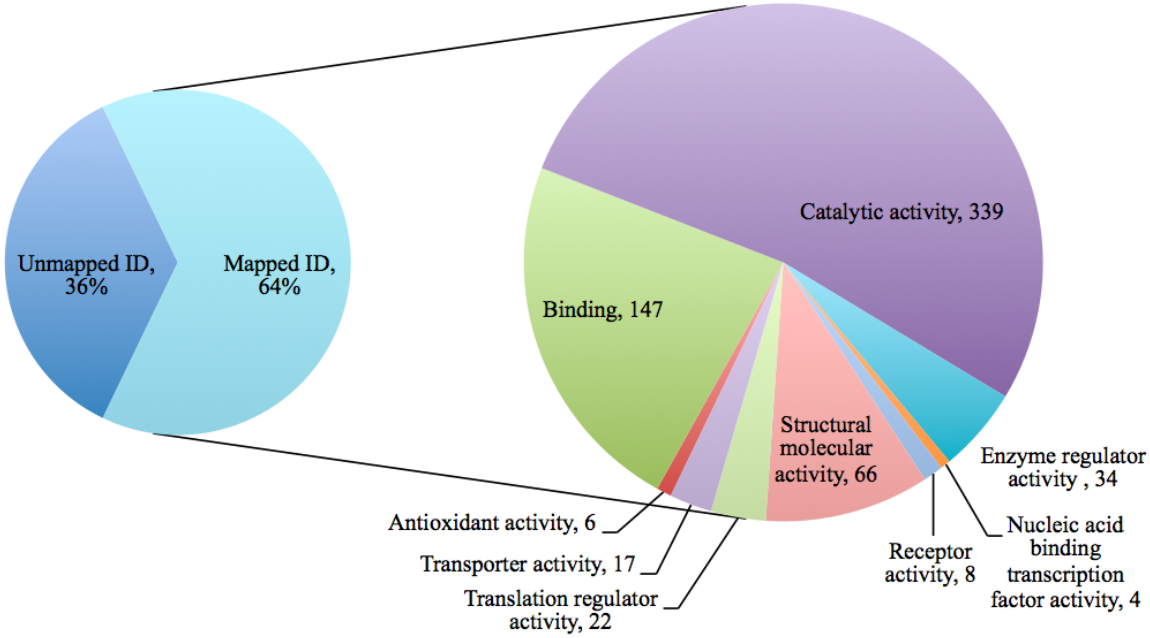


Figure 6.8 | Classification of Arabidopsis pollen coat proteins by 'molecular function'. The pie chart on the left indicates the percentage of mapped and unmapped proteomic IDs in the Gene Ontology database. The pie chart on the right represents the number of genes falling into each category of 'molecular function' as defined by the PANTHER GO-slim system (Mi *et al.*, 2013).

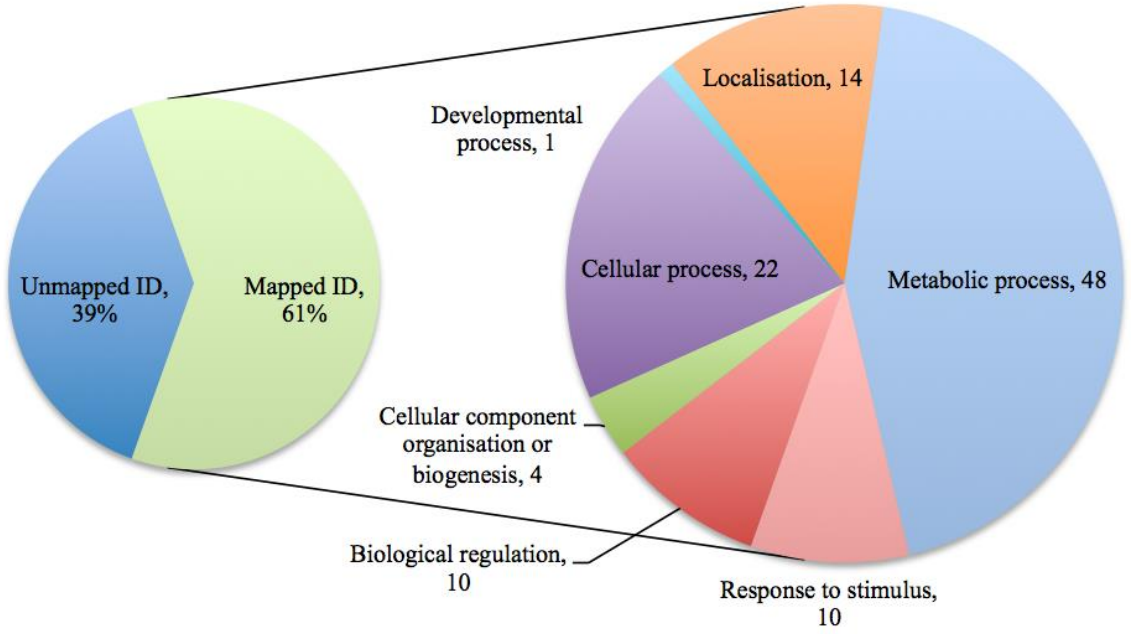


Figure 6.9 | Classification of Arabidopsis pollen coat proteins by 'biological process'. The pie chart on the left indicates the percentage of mapped and unmapped proteomic IDs in the Gene Ontology database. The pie chart on the right represents the number of genes falling into each category of 'biological process' as defined by the PANTHER GO-slim system (Mi *et al.*, 2013).

Table 6.2 | Protein molecular functions classified as 'catalytic activity' defined by PANTHER.

Gene IDs	PANTHER family/subfamily	PANTHER protein class
Q9SEU8	Thioredoxin M1, chloroplastic-related	oxidoreductase
Q8VZV6	Esterase/lipase/thioesterase family protein	phospholipase
Q8LA13	Dead-box ATP-dependent RNA helicase 11-related	RNA helicase
Q38879	Thioredoxin H2-related	oxidoreductase
Q3E6Q3	GTP-binding protein SAR1A-related	small GTPase
Q9C8W7	Subfamily not named	serine protease
Q9SNY3	GDP-mannose 4,6 dehydratase	dehydratase; epimerase/racemase; oxidoreductase
A8MS20	Subfamily not named	-
Q9FNE2	Glutaredoxin-2, mitochondrial	oxidoreductase
Q8VY11	DNA gyrase B3-related	DNA topoisomerase; enzyme modulator; isomerase
Q9FJR2	Cell wall / vacuolar inhibitor of fructosidase 2-related	-
P19172	Acidic endochitinase	glycosidase
Q9SRV5	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	methyltransferase
Q5MFV6	Pectinesterase 4-related	-
F4KCF2	NAD(P)-binding rossmann-fold superfamily protein	dehydrogenase; reductase
Q9FRL8	Chloride intracellular channel EXC-4	cytoskeletal protein; epimerase/racemase; reductase; signaling molecule; transferase; translation elongation factor
Q8H0X6	CPI-2, isoform A	cysteine protease inhibitor
Q8GXU8	1-acyl-SN-glycerol-3-phosphate acyltransferase	-
Q8L5T9	Cysteine proteinase inhibitor 2	cysteine protease inhibitor
Q56WF8	Serine carboxypeptidase-like 47-related	serine protease
P0DH91	ADP-ribosylation factor 2-B	small GTPase
P57752	Acyl-CoA-binding domain-containing protein 6	enzyme modulator; transfer/carrier protein
F4IBV4	Alpha-1,3/1,6-mannosyltransferase ALG2	carbohydrate phosphatase; glycosyltransferase; nucleotidyltransferase
O65351	Subtilisin-like protease SBT1.7-related	serine protease
B3H4Y0	Beta-galactosidase 1	galactosidase
Q8L7R3	1-acylglycerophosphocholine o-acyltransferase 1	acyltransferase; annexin; calmodulin
O23254	Serine hydroxymethyltransferase, cytosolic	methyltransferase
Q9T081	Anthocyanidin 3-o-glucoside 2'''-o-xylosyltransferase-related	-
Q42342	CytochromeB5 isoform E	oxidase
O04834	GTP-binding protein SAR1A-related	small GTPase
Q38935	Peptidyl-prolyl *Cis*-trans isomerase FKBP15-1-related	calcium-binding protein; chaperone; isomerase
Q8LFQ6	Glutaredoxin-C4	oxidoreductase
Q949V0	Subfamily not named	-
Q6Q151	Peptidyl-prolyl *Cis*-trans isomerase-like 4	isomerase
Q9SCV8	Beta-galactosidase 12-related	galactosidase
Q9SUL1	Subfamily not named	cysteine protease
Q9C9U3	Alpha-dioxygenase 2	oxygenase
P92976	Protein strictosidine synthase-like 1-related	lyase
F4K975	Motile sperm domain-containing protein 2	dehydrogenase
Q945Q1	Cysteine proteinase inhibitor 1	cysteine protease inhibitor
Q42406	Peptidyl-prolyl *Cis*-trans isomerase CYP18-4	isomerase
Q8L8T2	Glutaredoxin-C1	oxidoreductase
P32826	Serine carboxypeptidase-like 47-related	serine protease
Q9LNC1	Cathepsin-related	cysteine protease
Q8L7U0	Dehydrogenase/reductase SDR family protein 7-like	dehydrogenase; reductase
Q9FDW8	Cytochrome B5 isoform A	oxidase

Table 6.3 | Protein molecular functions classified as 'binding' defined by PANTHER.

Gene IDs	PANTHER family/subfamily	PANTHER protein class
P59230	60s ribosomal protein l10A-2-related	ribosomal protein
P19172	Acidic endochitinase	glycosidase
P57752	Acyl-CoA-binding domain-containing protein 6	enzyme modulator; transfer/carrier protein
P0DH91	ADP-ribosylation factor 2-B	small GTPase
Q94CG2	BET1-like snare 1-2	SNARE protein
Q945Q1	Cysteine proteinase inhibitor 1	cysteine protease inhibitor
Q8L5T9	Cysteine proteinase inhibitor 2	cysteine protease inhibitor
Q8H0X6	CPI-2, isoform A	cysteine protease inhibitor
Q8VY11	DNA gyrase B3-related	DNA topoisomerase; enzyme modulator; isomerase
Q9SF04	Gh01161P	signaling molecule
O04834	GTP-binding protein SAR1A-related	small GTPase
Q8L7R3	1-acylglycerophosphocholine o-acyltransferase 1	acyltransferase; annexin; calmodulin
Q93VB4	Programmed cell death protein 2	transcription cofactor
Q710E8	Origin recognition complex subunit 1	replication origin binding protein
P0C7R3	Pentatricopeptide (PPR) repeat-containing protein	RNA binding protein; serine/threonine protein kinase receptor; transporter
O23337	Subfamily not named	RNA binding protein; serine/threonine protein kinase receptor; transporter
Q38935	Peptidyl-prolyl *Cis*-trans isomerase FKBP15-1-related	calcium-binding protein; chaperone; isomerase
Q9LF54	Calcium-binding EF-hand family protein-related	calmodulin; enzyme modulator; signaling molecule
Q9SK39	Steroid-binding protein 3-related	receptor; signaling molecule
Q42449	Profilin-1-related	non-motor actin binding protein
Q3E6Q3	GTP-binding protein SAR1A-related	small GTPase
Q93YP0	UEV-1	-
Q8RUC6	Ubiquitin-NEDD8-like protein Rub2	ribosomal protein

Table 6.4 | Protein biological process classified as 'metabolic process' defined by PANTHER.

Gene IDs	PANTHER family/subfamily	PANTHER protein class
Q9SRV5	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	methyltransferase
P59230	60s ribosomal protein l10A-2-related	ribosomal protein
P19172	Acidic endochitinase	glycosidase
P57752	Acyl-CoA-binding domain-containing protein 6	enzyme modulator; transfer/carrier protein
Q8L7U0	Dehydrogenase/reductase SDR family protein 7-like	dehydrogenase; reductase
Q9SUL1	Subfamily not named	cysteine protease
Q949V0	Subfamily not named	-
Q945Q1	Cysteine proteinase inhibitor 1	cysteine protease inhibitor
Q8L5T9	Cysteine proteinase inhibitor 2	cysteine protease inhibitor
Q8H0X6	CPI-2, isoform A	cysteine protease inhibitor
Q9LNC1	Cathepsin-related	cysteine protease
Q9FDW8	Cytochrome B5 isoform A	oxidase
Q42342	Cytochrome B5 isoform E	oxidase
O23138	Cytochrome C-1	-
Q8LA13	Dead-box ATP-dependent RNA helicase 11-related	RNA helicase
Q8VY11	DNA gyrase B3-related	DNA topoisomerase; enzyme modulator; isomerase
Q9M0Y3	Equilibrative nucleotide transporter 2-related	transporter
Q944P0	Equilibrative nucleotide transporter 2-related	transporter
Q8VZV6	Esterase/lipase/thioesterase family protein	phospholipase
A8MS20	Subfamily not named	-
Q9SNY3	GDP-mannose 4,6 dehydratase	dehydratase; epimerase/racemase; oxidoreductase
Q8L8T2	Glutaredoxin-c1	oxidoreductase
Q9FNE2	Glutaredoxin-2, mitochondrial	oxidoreductase
Q8LFQ6	Glutaredoxin-C4	oxidoreductase
Q9FRL8	Chloride intracellular channel EXC-4	cytoskeletal protein; epimerase/racemase; reductase; signaling molecule; transferase; translation elongation factor
F4IBV4	Alpha-1,3/1,6-mannosyltransferase ALG2	carbohydrate phosphatase; glycosyltransferase; nucleotidyltransferase
Q710E8	Origin recognition complex subunit 1	replication origin binding protein
Q42406	Peptidyl-prolyl *Cis*-trans isomerase CYP18-4	isomerase
Q6Q151	Peptidyl-prolyl *Cis*-trans isomerase-like 4	isomerase
Q38935	Peptidyl-prolyl *Cis*-trans isomerase FKBP15-1-related	calcium-binding protein; chaperone; isomerase
Q9LXI4	Purple acid phosphatase 21-related	phosphatase
Q56WF8	Serine carboxypeptidase-like 47-related	serine protease
P32826	Serine carboxypeptidase-like 47-related	serine protease
O23254	Serine hydroxymethyltransferase, cytosolic	methyltransferase
P55852	Small ubiquitin-related modifier 1-related	-
Q9FLP6	Small ubiquitin-related modifier 1-related	-
Q38879	Thioredoxin H2-related	oxidoreductase
Q9SEU8	Thioredoxin M1, chloroplastic-related	oxidoreductase
Q93YP0	UEV-1	-
Q8RUC6	Ubiquitin-NEDD8-like protein Rub2	ribosomal protein

Table 6.5 | Protein biological process classified as 'cellular process' defined by PANTHER.

Gene IDs	PANTHER family/subfamily	PANTHER protein class
P42643	14-3-3-like protein GF14 Chi-related	chaperone
Q96300	14-3-3 protein epsilon	chaperone
Q01525	14-3-3-like protein GF14 Chi-related	chaperone
P42645	14-3-3 protein epsilon	chaperone
Q9FNP8	40s ribosomal protein S19-3	ribosomal protein
P0DH91	ADP-ribosylation factor 2-B	small GTPase
Q9SCK3	Subfamily not named	transporter
Q8LFH5	Bidirectional sugar transporter sweet 8	-
Q9SF04	Gh01161P	signaling molecule
O04834	GTP-binding protein SAR1A-related	small GTPase
Q9CAT6	Organic cation/carnitine transporter 1	carbohydrate transporter; cation transporter; transfer/carrier protein
Q42449	Profilin-1-related	non-motor actin binding protein
Q9FPJ4	RAS-related protein RABD2B-related	-
Q3E6Q3	GTP-binding protein SAR1A-related	small GTPase
Q9SRV5	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	methyltransferase
Q9SUL1	Subfamily not named	cysteine protease
Q949V0	Subfamily not named	-
Q9LNC1	Cathepsin-related	cysteine protease
O23138	Cytochrome C-1	-
Q8LA13	Dead-box ATP-dependent RNA helicase 11-related	RNA helicase
Q8VY11	DNA gyrase B3-related	DNA topoisomerase; enzyme modulator; isomerase
A8MS20	Subfamily not named	-
Q9FRL8	Chloride intracellular channel EXC-4	cytoskeletal protein; epimerase/racemase; reductase; signaling molecule; transferase; translation elongation factor
Q710E8	Origin recognition complex subunit 1	replication origin binding protein
Q42406	Peptidyl-prolyl *Cis*-trans isomerase CYP18-4	isomerase
Q6Q151	Peptidyl-prolyl *Cis*-trans isomerase-like 4	isomerase
Q38935	Peptidyl-prolyl *Cis*-trans isomerase FKBP15-1-related	calcium-binding protein; chaperone; isomerase
Q56WF8	Serine carboxypeptidase-like 47-related	serine protease
P32826	Serine carboxypeptidase-like 47-related	serine protease
P55852	Small ubiquitin-related modifier 1-related	-
Q9FLP6	Small ubiquitin-related modifier 1-related	-
Q38879	Thioredoxin H2-related	oxidoreductase
Q9SEU8	Thioredoxin M1, chloroplastic-related	oxidoreductase
Q93YP0	UEV-1	-

6.3 Discussion

Various components of pollen coat have been shown to play crucial roles during the early stages of pollen-stigma interaction (Doughty *et al.*, 1998; Fiebig *et al.*, 2000; Updegraff *et al.*, 2009). Although a number of studies spanning several decades have examined the protein profile of the pollen coat, the scarcity of the protein material and the historic paucity of information on gene and protein databases limited progress in uncovering the full complement of pollen coat protein components. Previously characterised proteins from pollen coat often represented abundant components, such as oleosin-like glycine-rich proteins (GRPs), lipases, protein kinases and caleosins (Mayfield *et al.*, 2001; Murphy, 2006; Wu *et al.*, 2015).

In this study, although the HPLC approach gave lower overall resolution and sensitivity, this method may be further utilised for assays that require native pollen coat proteins due to its ability to separate small proteins. LC-MS/MS analysis not only detected the previously discovered abundant proteins but also uncovered a dramatically wide range of proteins carried in the pollen coat of both Arabidopsis and *Brassica oleracea*. Previously characterised pollen coat proteins such as SCR and PCP-A1 were detected in the proteomic data reported here and importantly, the PCP-Bs, proteins identified as regulators of early post-pollination events (Chapter 3, Wang L *et al*., 2016), were also present in the data set. Thus the proteomic data presented in this study validated the presence of some PCP-Bs in the pollen coat, which previously were only predicted by a series of gene expression analyses (Chapter 3, Wang L *et al*., 2016). Our data sets also contained a large number of CRPs and small proteins <10 kDa that have previously not been directly detected in Arabidopsis pollen coat extracts (Table S6.1). Compared with the pollen coat protein profiles shown on by SDS-PAGE (Figure 6.1) and the previous published results achieved by Mayfield *et al*. (2001), the richness of the data sets obtained in this study demonstrated the high sensitivity of this approach. Only a small amount of pollen coat material was required and pre-fractionation of proteins by SDS-PAGE was avoided thus avoiding significant sample loss. LC-MS/MS superseded the other approaches in this study and thus no more analysis was performed by HPLC.

Amongst the gene families encoding the five largest CRP classes identified in the pollen coat from both *A. thaliana* and *B. oleracea*, some have been previously reported or

annotated but their specific functions remain unknown (Table 6.6). Our results demonstrate a potential pollen-related function for these genes. For example, gene families encoding two of the largest pollen coat CRP classes, the SCR-like proteins (SCRL) and PCP-A-like / low-molecular-weight cysteine-rich (PCPAL/ LCR) proteins, have been systematically identified as being homologous to the gene families that encode important factors relating to the self-incompatibility system in *Brassica* (Vanoosthuyse *et al.*, 2001). Expression analyses by RT-PCR have indicated that several members of the SCRL and PCPAL protein families are expressed in flower buds (Vanoosthuyse *et al.*, 2001), and this is partly supported by our proteomic data set for *B. oleracea* pollen coat.

Table 6.6 | Descriptions of annotated proteins in five largest CRP classes in UniProtKB.

PCP-A/DEFL		
Accession	**Gene code**	**Description**
P82746	*At1g28335*	Defensin-like protein 153/LCR31
Q9C947	*At1g61070*	Defensin-like protein 5
Q9ZUL8	*At2g02140*	Defensin-like protein 10
Q8S8H9	*At2g15535*	Defensin-like protein 144
Q2V466	*At2g22941*	Putative defensin-like protein 191
P82729	*At2g25344*	Putative defensin-like protein 137/LCR14
Q8S8H3	*At2g28355*	Defensin-like protein 149 /LCR5
P82747	*At2g28405*	Putative defensin-like protein 150/LCR32
O22866	*At2g43520*	Defensin-like protein 193
O22867	*At2g43530*	Defensin-like protein 194
P82719	*At3g25265*	Putative defensin-like protein 148/LCR4
P82749	*At4g09984*	Putative defensin-like protein 142/LCR34
Q2V3K0	*At4g10603*	Putative defensin-like protein 169
Q9T0E3	*At4g11760*	Defensin-like protein 151/LCR17
P82739	*At4g29305*	Defensin-like protein 159/LCR25
P82716	*At5g48543*	Defensin-like protein 147/LCR1
P82748	*At3g43083*	Putative defensin-like protein 133/LCR33
Q2V2W7	*At5g60615*	Putative defensin-like protein 274
PCP-B/PCP-BL		
Accession	**Gene code**	**Description**
A8MQY8	*At2g16505*	EMBRYO SURROUNDING FACTOR 1-like protein 9 (ESFL9)
A8MR88	*At2g16535*	EMBRYO SURROUNDING FACTOR 1-like protein 8 (ESFL8)
Q1G3R6	*At2g41415*	EMBRYO SURROUNDING FACTOR 1-like protein 4 (ESFL4)
Q1PDG8	*At5g61605*	EMBRYO SURROUNDING FACTOR 1-like protein 10 (ESFL10)
SCRL		
Accession	**Gene code**	**Description**
P82644	*At4g32714*	Putative defensin-like protein 231/SCRL25
P82643	*At4g32717*	Putative defensin-like protein 230/SCRL24
P82622	*At1g08695*	Putative defensin-like protein 228/SCRL3
P82639	*At4g10115*	Putative defensin-like protein 236/SCRL20
P82635	*At2g06983*	Putative defensin-like protein 238/SCRL16
P82642	*At4g14785*	Defensin-like protein 232/SCRL23
P82646	*At5g45875*	Defensin-like protein 229/SCRL27
P82621	*At1g65113*	Defensin-like protein 226/SCRL2
P82641	*At4g33465*	Putative defensin-like protein 233/SCRL22
P82638	*At3g27503*	Defensin-like protein 241/SCRL19
LTP/LTPL		
Accession	**Gene code**	**Description**
Q2V3C1	*At4g33355*	Non-specific lipid-transfer protein 11(nsLTP11)
Q8GT78	*At5g62080*	Protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
Q00762	*At5g07230*	Tapetum-specific protein A9
Q9LLR6	*At5g59310*	Non-specific lipid-transfer protein 4 (nsLTP4)
A8MQR9	*At3g11825*	Protease inhibitor/seed storage/LTP family protein
CRRSP		
Accession	**Gene code**	**Description**
Q9LRK8	*At3g21990*	Cysteine-rich repeat secretory protein 29 (CRRSP29)
Q9LRK4	*At3g22030*	Putative cysteine-rich repeat secretory protein 33 (CRRSP33)
Q9M2I5	*At3g58310*	Putative cysteine-rich repeat secretory protein 61 (CRRSP61)
Q9LRL4	*At3g21960*	Putative cysteine-rich repeat secretory protein 24 (CRRSP24)
Q9LRK2	*At3g22050*	Putative cysteine-rich repeat secretory protein 35 (CRRSP35)
Q9LRL1	*At3g21970*	Cysteine-rich repeat secretory protein 26 (CRRSP26)

CRRSPs have been identified in Arabidopsis as a non-kinase protein family that possesses a motif consisting of two copies of a cysteine-rich region (Table 6.1) (Chen, 2001). This cysteine-rich repeat (CRR) motif is also possessed by a kinase protein superfamily CRR RLKs (CRKs). Later studies revealed that several CRRSPs were located on the plasma membrane at plasmodesmata (PD) and were renamed as PD-located proteins (PDLPs) (Thomas *et al.*, 2008). Functional studies implied that PDLPs promote cell-to-cell movement of plant viruses (Amari *et al.*, 2010). Their receptor-like properties were found to be important for assembling viral movement proteins (MP) into tubules which facilitate viral movements between cells through plasmodesmata (Amari *et al.*, 2010). However, no previous studies have shown that CRRSPs are located in the pollen coat or that they possess any function related to plant sexual reproduction.

Non-specific lipid-transfer proteins (nsLTPs) have been identified in a wild range of plant families from monocot, eudicot and even non-flowering plant lineages (reviewed in Liu *et al.*, 2015). A large number of studies have uncovered functions of nsLTPs relating to plant reproduction. For example, several genes, such as OsC6 in *Oryza sativa (Zhang et al., 2010),* E2 PLTP in *Brassica napus* (Foster *et al.*, 1992), CaMF2 in *Capsicum annuum* (Chen *et al.*, 2011), A9 in *Arabidopsis thaliana* (Paul *et al.*, 1992; Ariizumi *et al.*, 2002) have been identified as being exclusively expressed in anthers and especially in the tapetum during the early stages of anther development. nsLTPs have also been identified as being involved in the generation of the pollen exine, formation of anther epidermal cells, adhesion of pollen tubes to the stigmatic transmitting tract and pollen tube growth (Park *et al.*, 2000; Park & Lord, 2003; Jung *et al.*, 2006; Chae *et al.*, 2009; Chae *et al.*, 2010; Huang *et al.*, 2013). The detection of multiple nsLTPs from *A. thaliana* and *B. oleracea* pollen coat in this study corresponds well with the aforementioned studies, and provides further evidence for the important roles played by this protein family during plant reproductive development and potentially pollen-stigma interactions. In summary, the pollen coat proteomic data obtained in this project revealed large numbers of CRPs with functions that have yet to be elucidated and thus provides a solid platform from which to launch future functional analyses of these proteins.

A recent phylostratigraphic profiling of rice and Arabidopsis genes revealed a series of young protein-coding genes (Cui *et al.*, 2015). The comparison of genome features of old and young genes in both species indicated that young genes possess fewer exons and

encode smaller proteins. Moreover, though young genes are much less well functionally annotated, gene ontology (GO) enrichment analysis suggested that young genes are likely involved in defense and reproduction whereas old genes are related to primary metabolism (Cui *et al.*, 2015). Furthermore, GO cellular localisation predictions demonstrated that proteins encoded by young genes are more likely to be targeted to the extracellular environment, endomembrane system or be anchored to a membrane (Cui *et al.*, 2015). These features of young genes and their encoded proteins correspond to our proteomic analysis of the Arabidopsis and *Brassica oleracea* pollen coat as well as previous studies on pollen coat CRPs (Chapter 3, Marshall *et al.*, 2011). Interestingly, the quantification of transcriptome age suggested that in evolutionary terms the youngest transcriptome is expressed in pollen (Cui *et al.*, 2015). In addition, Cui *et al.* (2015) also observed that young genes in Arabidopsis and rice are not likely to be localized to block duplicated regions on chromosomes, which corresponds to the pattern of gene duplication for the *PCP-B*-like (*PCPBL*) genes (Chapter 5).

As mentioned in Chapter 1 (section 1.1.2), the phenomenon of adaptive divergence has been commonly observed amongst genes involved in sexual reproduction and reproductive isolation (Swanson & Vacquier, 2002). Molecular evolutionary studies on CRP families including SCRL, PCPAL, nsLTP and PCPBL revealed evidence of positive selection force acting on these proteins (Vanoosthuyse *et al.*, 2001; Jang *et al.*, 2008) (Chapter 5). The highly duplicated and diversified genes encoding pollen coat CRPs may provide the evolutionary raw material and driving force for the formation of reproductive barriers and speciation. Our proteomic analysis revealed a ‘reservoir’ of potential protein regulators in plant reproductive signalling and provides the basis for research on pollen coat function in the future.

Chapter 7 General discussion and Conclusions

The establishment of reproductive barriers through mechanisms that govern pollen-stigma compatibility is a key feature of intraspecific and interspecific pollination amongst angiosperms. Indeed, the ability to control mating partners prezygotically has been an important evolutionary component of angiosperm diversification and speciation (reviewed in Rieseberg & Willis, 2007). Although over the last several decades, studies focusing on the mechanisms of self-incompatibility (SI) in angiosperms have made significant inroads into our understanding of this phenomenon, the molecular basis of self-recognition culminating in compatibility is still poorly understood.

In plant cell-cell communication, ligand-receptor pairs have been found to play central roles in a variety of processes including meristem development, cell expension, innate immunity, pathogene response, symbiotic signalling and pollen-pistil interactions (reviewed in Matsubayashi, 2003). However identifying the ligand-receptor pairs is chanlenging. Genome of *Arabidopsis thaliana* encodes more than 600 receptor-like kinases (RLKs) and over 1000 putative secreted ligands (Shiu & Bleecker, 2001; Lease & Walker, 2006). Only very few ligand-receptor pairs have been identified (See Chapter 4, 4.1) (Endo *et al*., 2014). Although in recent years, some of the 'orphan' ligands and receptors were paired such as RALF and FERONIA (Haruta *et al*., 2014), LURE and MDIS1/MIK (Wang T *et al*., 2016), the ligands of most RLKs remain unclear.

In recent years, a broad family of secreted cysteine-rich proteins (CRPs) have been found to have a range of roles in plant reproduction and several of them have now been shown to act as pollen coat signalling ligands that interact with stigmatic receptors (see Chapter 1, 1.3.3.1) (Doughty *et al.*, 1998; Schopfer *et al.*, 1999; Takayama *et al.*, 2000; Takayama *et al.*, 2001). Thus members of the CRP family of proteins might act as ligands for many 'orphan' RLKs in plants. Although this superfamily of proteins can be divided into many classes due to their structural differences and the number and pattern of cysteine residues, some common features are shared and can be utilised in the identification of new members. Nearly all the CRPs possess an N-terminal signal peptide and a mature region with conserved cysteine residues (see Chapter 1, Figure 1.6). CRP-encoding genes appear to have single or double exons with a conserved intron position and are commonly found to be clustered on the genome. As proposed by Silverstein *et al.* (Silverstein *et al.*, 2007),

CRP classes in plants may evolve by rearrangements of cysteine motifs. Interestingly, many members of the CRP superfamily also possess proline-rich peptides (PRPs) and glycine-rich peptides (GRPs) at the N or C- terminal regions or even at internal positions (Silverstein *et al*., 2007) (Table 7.1). Additionally, internal duplication and cysteine-rich motif fusion have been identified in almost every CRP class (Silverstein *et al*., 2007). Such observations indicated that these 'chimeric proteins' might hold the potential to be cleaved to play multiple roles in plant life, raising the possibility of an origin of gamete recognition CRPs from proteins involved in plant-microbe interactions. The detection of a large number (more than 300) of defensin-like genes in Arabidopsis provided a possible evolutionary and functional link between CRPs in plant immune and reproductive systems.

Table 7.1 | CRP fusions with PRP, GRP and other CRPs. SP, signal peptide, CRP, cysteine-rich peptide, GRP, glycine-rich peptide, CRP', a CRP with distinct cysteine pattern, CRP*, a CRP with interrupted domain. Table adapted from Silverstein *et al*., 2007.

Observed pattern	Number of subgroups affected	Number of unique suqences
SP-CRP-PRP	33	1269
SP-CRP-CRP	23	329
SP-PRP-CRP	18	88
SP-CRP-CRP'	16	65
SP-CRP*-GRP-CRP*	13	361
SP-GRP-CRP	6	72
SP-CRP*-PRP-CRP*	5	76
SP-PRP-CRP-PRP	4	70
SP-CRP-GRP	3	14

Interestingly, by comparing the stigmatic responses to those initiated during plant defence, it is not difficult to identify similarities between signalling responses elicited by both the pollen-stigma interaction and plant-microbe interaction. PCP-A1, a pollen coat protein found to interact with a stigmatic surface *S* locus glycoprotein (SLG), possesses a conserved cysteine pattern that is highly similar to that of plant defensins (Doughty *et al.*, 1998). Dynamics of cytosolic free calcium concentration ($[Ca^{2+}]_{cyt}$) were observed in both early compatible and incompatible pollination processes (Iwano *et al.*, 2004; Wu *et al*., 2011) and plant defence-signalling pathways (reviewed in Lecourieux *et al.*, 2006). Signals mediated by the redox network such as reactive oxygen species (ROS) and reactive nitrogen species (RNS) in plants play important roles not only in the responses to stress and pathogens (Figure 7.1) (reviewed in Baxter *et al.*, 2014; Petrov *et al.*, 2015; Stael *et al*., 2015), but intriguingly, also in sexual plant reproduction such as SI system in *Papaver*

(reviewed in Traverso *et al.*, 2013; Eaves *et al.*, 2014) (Figure 7.2). The sharing of signalling molecules between pathways mediating pollen-pistil recognition and responses to pathogens indicates a possible linked evolutionary history for the plant immune and reproductive systems: the reproductive recognition molecules might have evolved from components of the plant immune system whilst the downstream signalling networks have been maintained.

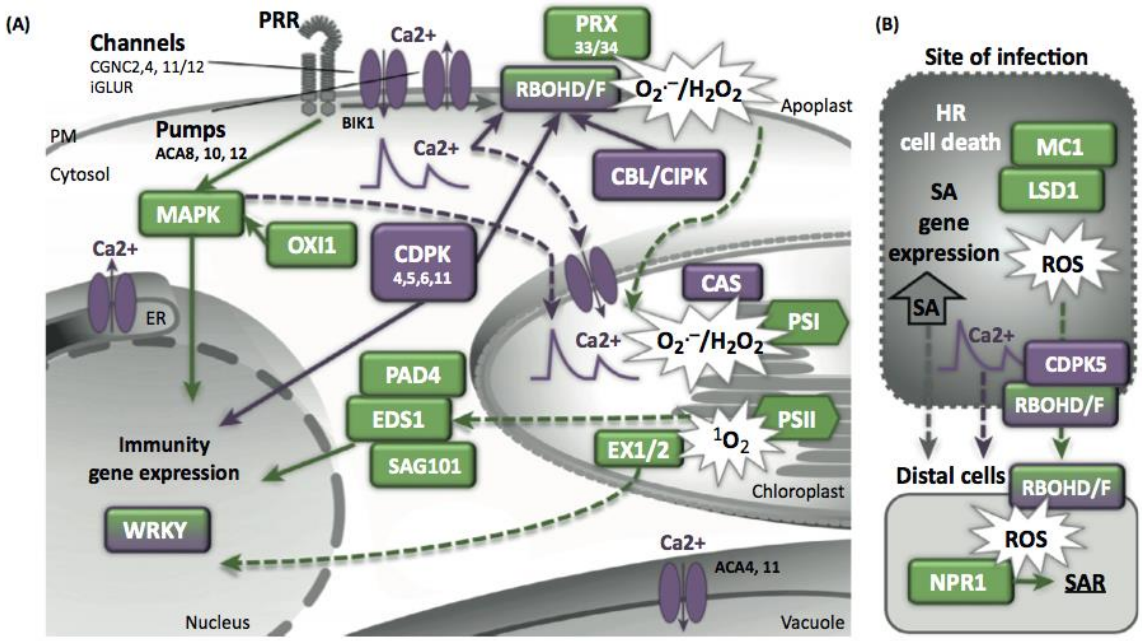


Figure 7.1 | Calcium signalling pathways and reactive oxygen species (ROS) in plant innate immunity. (A) The perception of pathogen associated molecular pattern (PAMP) by pattern recognition receptors (PRRs) triggers rapid response, signalling the action of calcium (Ca^{2+}) channels and transporters that result in a cytosolic Ca^{2+} flux. Activated by Ca^{2+}, a signalling cascade involving calcium-dependent protein kinases (CDPKs) and a mitogen-activated protein kinase (MAPK) triggers expression of immunity gene expression in the nucleus such as WRKY transcription factors. MAPKs are regulated by the ROS sensory kinase oxidative signal-inducible 1 (OXI1). Ca^{2+} influx, CDPKs phosphorylated by Botrytis-induced kinase 1 (BIK1) and calcineurin B-like protein (CBL)/ CBL-interacting protein kinase (CIPK) modules enhance the activity of respiratory burst oxidase homologs (RBOHs) D and/or F (RBOHD/F localised on the plasma membrane, which produces apoplastic ROS. Peroxidases 33 and 34 (PRX33/34) also contribute to the generation of apoplastic ROS for the oxidative burst. The rapid generation (within 20 min) of Ca^{2+} flux in the chloroplast is regulated by the thylakoid associated calcium-sensing protein (CAS). Downstream retrograde signaling to the nucleus might be mediated by the ROS $^1O_2$ generated by photosystem II (PSII) and $O_2^{\cdot-}$ generated by photosystem I (PSI). Downstream of $^1O_2$ Executer 1 and 2 (EX1/2) alter nuclear gene expression. The central immune regulator enhanced disease susceptibility 1 (EDS1) has been proposed to be the downstream target of chloroplastic $^1O_2$ and interacts with phytoalexin deficient 4 (PAD4) and senescence-associated gene 101 (SAG101) to form a heterodimer to alter nuclear immunity gene expression. (B) The ROS sensory protein, lesion simulating disease 1 (LSD1) mediates a later response of plant cells to pathogens, which was hypothesised to be an inhibitor of spreading of cell death lesions. This process is mediated by metacaspase 1 (MC1) during hypersensitive response (HR)-type cell death. Upregulated immunity gene expression leads to an increase in level of the immune hormone salicylic acid (SA) in the chloroplast. Spreading of the SA signal, calcium flux and ROS signal could target the ROS sensory protein, nonexpressor of pathogenesis-related genes 1 (NPR1) in *Arabidopsis* to activate systemic acquired resistance (SAR) in distal cells at the site of infection. Proteins related to Ca^{2+} signaling are shown in purple. Proteins related to ROS signaling are shown in green. Hypothetical connections are shown as dashed lines. CGNC, Ca^{2+}-permeable cyclic nucleotide-gated channel; ER, endoplasmic reticulum; iGluR, ionotropic glutamate receptor-like channels; PM, plasma membrane; ACA4, 8, 10, 11, 12, autoinhibited calcium-ATPase 4, 8, 10, 11, 12. Figure adapted from Stael *et al.*, 2015.

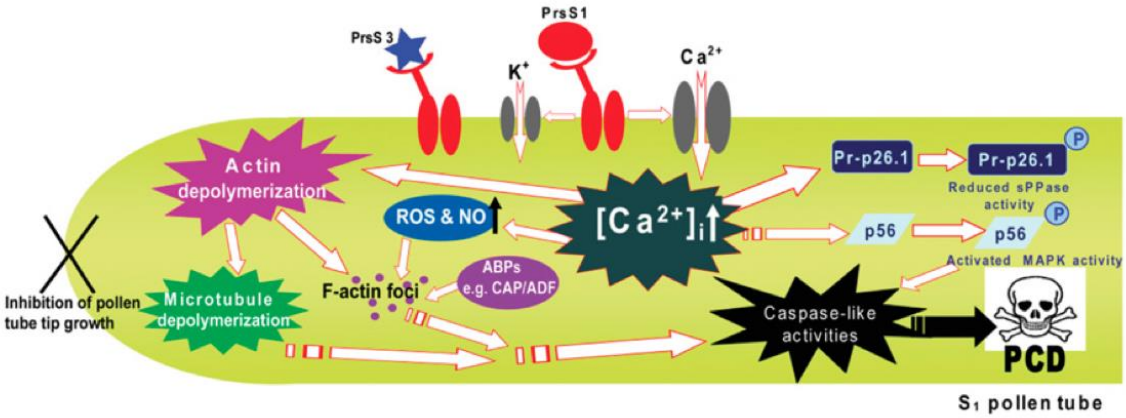


Figure 7.2 | Integration of the signals and targets of the *Papaver* self-incompatibility signalling network. Interaction of PrsS and PrpS proteins triggers K^+ and Ca^{2+} influx. Increases in the concentration of cytosolic Ca^{2+} triggers a downstream signalling network mediated by multiple components, which results in the inhibition of pollen tube growth and PCD of incompatible pollen tube. The phosphorylated sPPase, Pr-p26 and Ca^{2+} result in the inhibition and reduction of biosynthetic capacity. The F-actin cytoskeleton is rapidly depolarised and forms punctate foci, which contributes to the pollen tube growth inhibition. The microtubule cytoskeleton is also rapidly depolymerized, which results in PCD triggered by caspase-like activities and phosphorylated MAPK and p56. ROS and NO act as upstream signalling molecules for the formation of actin foci and activation of caspase-3-like proteins and eventually PCD. This signalling network blocks pollen tube growth and thus inhibits the fertilisation in an incompatible pollen-pistil interaction. ABP, actin-binding protein; CAP, cyclase-associated protein. Figure adaptive from Eaves *et al*., 2014.

The link between the plant immune and reproductive system is also supported by the fact the disease resistant (*R*) genes playing important roles in the evolution of hybrid invalidity. Hybrid invalidity acts as postzygotic reproductive barriers and is characterised by hybrid necrosis, weakness or tumour growth in intra- or inter-specific hybrids, features which are commonly associated with environmental stress and pathogen attack (reviewed in Bomblies & Weigel, 2007). Enhanced disease resistance has been observed to be associated with mild hybrid necrosis. For example, resistance breeding programmes of *Solanum lycopersicon* (domestic tomato) and *Solanum tuberosum* (domastic potato) produced plants with high virus-resistance and antonecrosis, a spontaneous appearance of lesions that resemble pathogen response yet without pathogen attack (Valkonen & Watanabe, 1999). In wheat, the correspondence of resistance to rust fungus and spot blotch to a non-lethal leaf-tip necrosis has been utilised by breeders as a visible marker for the resistance genotype (Joshi *et al*., 2004). These correlations indicated that the selection pressure from pathogen attack or human breeding programmes for disease resistance might be directly or indirectly affecting the evolution of deleterious linked-loci interactions that could potentially form reproductive barriers in plants.

The tightly linked loci of hybrid necrosis and disease resistance suggested that genes which occur in large clusters are more likely to be responsible (Bomblies & Weigel, 2007). A recent study showed the interaction of an *R* gene (*R3*) that belongs to Toll/interleukin-1 receptor-nucleotide binding-leucine rich repeat (TNL) *RPP1 (Recognition of Peronospora parasitica1)-like genes* in *A.thaliana (Ler)* and alleles of the *receptor-like kinase (RLK) Strubbelig Receptor Family 3 (SRF3)* in Central Asian strains *Kas-2* or *Kond* were responsible for temperature-dependent autoimmunity and low growth and reproductive fitness of the hybrids (Alcazar *et al*., 2014). Interestingly, cysteine-rich proteins (CRPs) identified in Arabidopsis and *Brassica oleracea* were found to be encoded by genes that frequently appear in large clusters such as the SCR-like proteins (SCRL), PCP-A-like / low-molecular-weight cysteine-rich (PCPAL/ LCR) proteins and cysteine-rich repeat secretory proteins CRRSP (Chapter 6). The presence of these proteins on the pollen surface and the similarity between PCPAL/ LCR and plant defensins suggest a possibility of the linkage between the pathogen or disease resistance of pollen grains and the formation of early post-pollination prezygotic barriers.

The molecular evolutionary study provided evidence of positive selection on the codon sites of putative *PCP-B* orthologous genes, which is a common feature of reproductive proteins and 'speciation' genes. The correspondence of the codon sites under positive selection with amino acid sites that have positive electrostatic potential on the protein surface also opened up the intriguing possibility that these 'hotspots' may be important binding sites for AtPCP-Bs with their stigmatic targets. Based on the model for the molecular evolution of the mechanism that operates in species-specific system in abalone (Swanson & Vacquier, 2002), the evolution of stigmatic receptors might also have been providing the selection force for the evolution of PCP-Bs. Although nothing is known about the evolutionary biology of the possible direct stigmatic receptors, a model can be proposed to explain the current observations during the early stages of pollen-stigma interactions in the Brassicaceae. As the first checkpoint of pollen-stigma compatibility, pollen hydration is regulated by cell-cell communication mediated by pollen coat ligands and stigmatic receptors. Similar to the lysin-VERL interaction model in abalone (Swanson & Vacquier, 2002, Chapter 1, 1.1.2, Figure 1.2), stigmatic receptors may also be evolving neutrally due to possible gene conversion among gene copies or tandem repeat regions of the receptors. The changes to individual receptors or repeat regions of stigmatic receptors might cause lower affinity with pollen ligands whilst still allowing for pollen hydration.

The hydration of pollen carrying ligands with lower receptor-binding affinity could be impaired, which could cause a delay in pollen tube growth. Thus changes to stigmatic receptors applies selective pressure onto genes encoding pollen coat ligands and results in their coevolution. Plant speciation results from the formation of multiple reproductive barriers including pre-pollination barriers, post-pollination prezygotic barriers (such as pollen hydration and pollen tube guidance) and postzygotic barriers (reviewed in Rieseberg & Willis, 2007). Thus species-specificity of factors mediating pollen hydration may not be strong enough to lead to complete reproductive isolation. Indeed pollen hydration does occur between closely related species (Hiscock & Dickinson, 1993), whose reproductive isolation is dominantly performed by downstream prezygotic barriers or postzygotic barriers. Intriguingly, interspecific unilateral incompatibility (UI) has been frequently reported during the last several decades (reviewed in Onus & Pickersgill, 2004; Covey *et al.*, 2010; Kitashiba & Nasrallah, 2014). In the Brassicaceae, these unequal post-pollination, prezygotic reproductive barriers preventing hybridisation are commonly observed when a SI species is the female parent and a closely related SC species is the male parent, which is termed as the 'SI x SC rule' (Hiscock & Dickinson, 1993). The molecular mechanism of UI in the Solanaceae has been thoroughly studied and reveals a strong link between UI and the SI system in this family (Li & Chetelat, 2014; Tovar-Mendez *et al.*, 2014; Li & Chetelat, 2015). In the Brassicaceae, although some interspecific incompatibility analyses have implicated links between SI and interspecific-incompatibility (Udagawa *et al.*, 2010), no molecular evidence has been found to support such a model. Thus, if the abalone lysin-VERL evolution model does apply to pollen ligands and their receptors in the Brassicaceae, the phenomenon of UI might be explained by this theory (Figure 7.3): here, shortly after the isolation of two populations, the divergence of pollen hydration factors accumulates however not enough to provide complete species-specificity and thus interspecific pollen hydration can occur. Once the transformation from SI to SC occurs in one species, the selective pressure from the ever-changing stigmatic receptors is relaxed on pollen ligands due to the less competitive environment on papillar cells, where barely any pollen from other individuals and species compete with self-pollen. This lack of constraint allows the pollen ligand-coding genes to rapidly accumulate mutations – at this point pollen hydration can still occur, however a lower affinity exists between pollen ligands and their stigmatic receptors. Consequently, the rapidly diverged pollen ligands present in SC species cannot be recognised by the stigmas of closely related SI species and thus results in the 'SI x SC rule' that has been observed in Brassicaceae. However, exceptions to the 'SI x SC rule' have been observed recently in the Doughty lab (Lian Fan - unpublished data) between

populations of *A. thaliana* and *A. lyrata*, suggesting that the 'SI x SC rule' may not be robust between recently diverged species. Here substantial variation has been observed within SI populations.

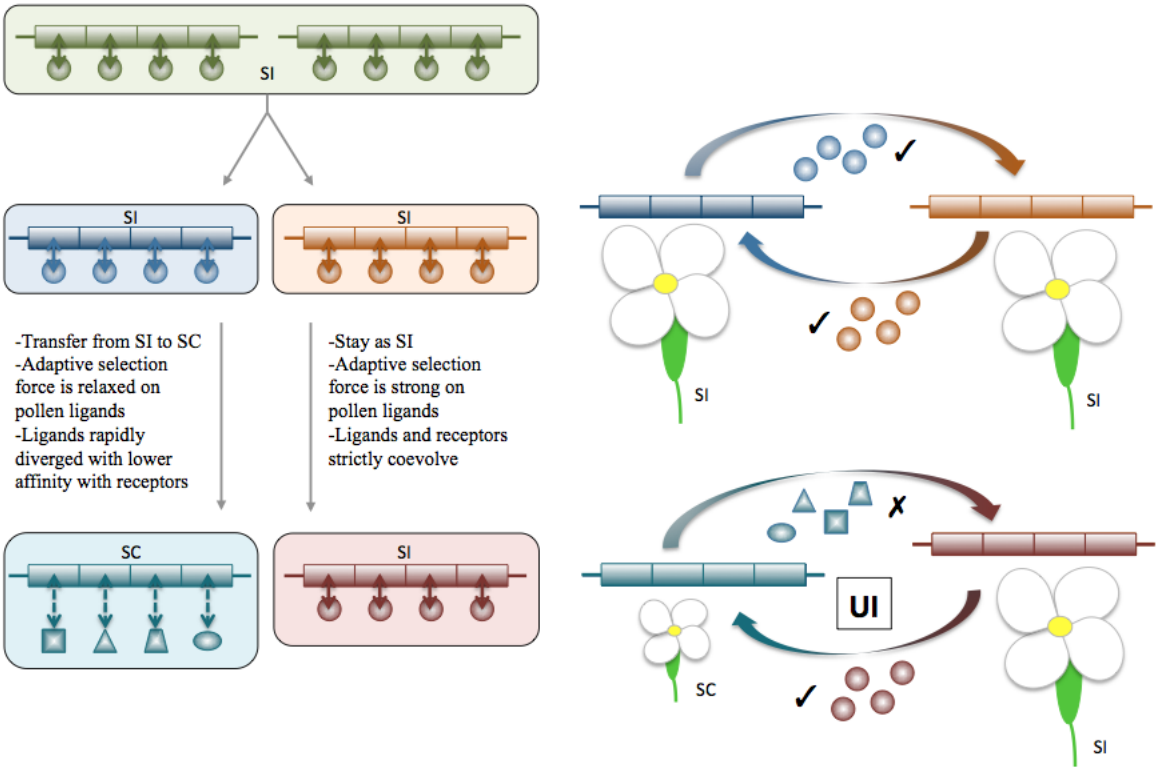


Figure 7.3 | Differences in the coevolution between pollen ligands and stigmatic receptors might explain the phenomenon of interspecific unilateral incompatibility (UI). Stigmatic receptors are represented as bars and pollen ligands are represented as circles in SI species and diverged shapes in SC species. Once speciation occurs between two populations, the receptors evolve but still remain similar. In SI species, pollen ligands coevolve with receptors to maintain strong affinity, whereas in SC species, the selective pressure is relaxed leading to rapid diversification of ligand-coding genes and lowered affinity between ligands and receptors. Thus the 'SI x SC rule' is observed between members of the Brassicaceae.

Although the molecular evolutionary study in this project requires more thorough analyses to further support any hypothesis, the methods performed here could be utilised for future research in intra and interspecific pollen-stigma compatibility. For example, the molecular evolutionary analysis utilising the maximum likelihood method could be used for the study of coevolution between pollen ligands and their potential stigmatic receptors when they are identified. The proteomic analyses of the Arabidopsis and *Brassica oleracea* pollen coats revealed a dramatically diverse array of proteins, with most having no known function. These data sets have provided a precious collection of proteins that will be important for future studies aimed at identifying factors that mediate the pollen-stigma interaction. In particular they will be important for research on unilateral incompatibility (UI) in the

Brassicaceae as these data sets likely contain crucial regulatory factors that mediate pollen recognition and acceptance by interspecific stigmas. Further, similar phylogenetic analyses and the study of selection forces could be applied in future studies on the other classes of pollen coat CRPs identified in this project that may have roles in reproductive signalling.

In summary, this PhD project provided exciting new insights into the molecular basis of the regulation of early stages of pollination in Arabidopsis and thus pollen-stigma compatibility. Future work needs to focus on further loss-of-function studies by creating an Arabidopsis mutant line carrying knockouts for all four AtPCP-Bs, as well as, importantly, continuing the search for their stigmatic targets. The future characterisation of novel pollen coat proteins will also shed new light on understanding the mechanisms of interspecies and intraspecies compatibility in sexual plant reproduction and angiosperm diversification.

References

Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, Ma H, Palmer JD, Rounsley S, Sankoff D, Schuster SC, *et al*. 2013. The Amborella genome and the evolution of flowering plants. *Science* **342**(6165): 1467-1476.

Alcazar R, von Reth M, Bautor J, Chae E, Weigel D, Koornneef M, Parker JE. 2014. Analysis of a plant complex resistance gene locus underlying immune-related hybrid incompatibility and its occurrence in nature. *Plos Genetics* **10**(12): e1004848.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen HM, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, *et al*. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**(5633): 653-657.

Alunni B, Kevei Z, Redondo-Nieto M, Kondorosi A, Mergaert P, Kondorosi E. 2007. Genomic organization and evolutionary insights on *GRP* and *NCR* genes, two large nodule-specific gene families in *Medicago truncatula*. *Molecular Plant-Microbe Interactions* **20**(9): 1138-1148.

Amari K, Boutant E, Hofmann C, Schmitt-Keichinger C, Fernandez-Calvino L, Didier P, Lerich A, Mutterer J, Thomas CL, Heinlein M, *et al*. 2010. A family of plasmodesmal proteins with receptor-like properties for plant viral movement proteins. *Plos Pathogens* **6**(9): e1001119.

Amien S, Kliwer I, Marton ML, Debener T, Geiger D, Becker D, Dresselhaus T. 2010. Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *Plos Biology* **8**(6): e1000388.

Anderson MA, Cornish EC, Mau SL, Williams EG, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche PJ, *et al*. 1986. Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. *Nature* **321**(6065): 38-44.

Ariizumi T, Amagai M, Shibata D, Hatakeyama K, Watanabe M, Toriyama K. 2002. Comparative study of promoter activity of three anther-specific genes encoding lipid transfer protein, xyloglucan endotransglucosylase/hydrolase and polygalacturonase in transgenic *Arabidopsis thaliana*. *Plant Cell Reports* **21**(1): 90-96.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**(2): 195-201.

Arnold ML, Bulger MR, Burke JM, Hempel AL, Williams JH. 1999. Natural hybridization: How low can you go and still be important? *Ecology* **80**(2): 371-381.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**(1): 25-29.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**: W202-W208.

Balakirev ES, Ayala FJ. 2003. Pseudogenes: Are they "Junk" or functional DNA? *Annual Review of Genetics* **37**: 123-151.

Balandin M, Royo J, Gomez E, Muniz L, Molina A, Hueros G. 2005. A protective role for the embryo surrounding region of the maize endosperm, as evidenced by the characterisation of *ZmESR-6*, a defensin gene specifically expressed in this region. *Plant Molecular Biology* **58**(2): 269-282.

Barrett SCH. 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics* **3**(4): 274-284.

Barrett SCH. 2010. Understanding plant reproductive diversity. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**(1537): 99-109.
Baxter A, Mittler R, Suzuki N. 2014. ROS as key players in plant stress signalling. *Journal of Experimental Botany* **65**(5): 1229-1240.
Beale KM, Leydon AR, Johnson MA. 2012. Gamete fusion is required to block multiple pollen pubes from entering an Arabidopsis ovule. *Current Biology* **22**(12): 1090-1094.
Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **107**(43): 18724-18728.
Bergsten J. 2005. A review of long-branch attraction. *Cladistics* **21**(2): 163-193.
Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, *et al*. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**(W1): W252-W258.
Blake JA, Christie KR, Dolan ME, Drabkin HJ, Hill DP, Ni L, Sitnikov D, Burgess S, Buza T, Gresham C, *et al*. 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Research* **43**(D1): D1049-D1056.
Blanc G, Barakat A, Guyot R, Cooke R, Delseny I. 2000. Extensive duplication and reshuffling in the arabidopsis genome. *Plant Cell* **12**(7): 1093-1101.
Blanc G, Hokamp K, Wolfe KH. 2003. A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Research* **13**(2): 137-144.
Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **103**(15): 5869-5874.
Bomblies K, Weigel D. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* **8**(5): 382-393.
Bornhorst JA, Falke JJ. 2000. Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology* **326**: 245-254.
Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. 2009. Yeast-two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences* **10**(6): 2763-2788.
Bruix M, Jimenez MA, Santoro J, Gonzalez C, Colilla FJ, Mendez E, Rico M. 1993. Solution structure of gamma-1-H and gamma-1-P thionins from barley and wheat endosperm determined by H1-NMR - a structural motif common to toxic arthropod proteins. *Biochemistry* **32**(2): 715-724.
Byng JW, Chase MW, Christenhusz MJM, Fay MF, Judd WS, Mabberley DJ, Sennikov AN, Soltis DE, Soltis PS, Stevens PF, *et al*. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**(1): 1-20.
Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology* **4**: 10.
Chae K, Gonong BJ, Kim S-C, Kieslich CA, Morikis D, Balasubramanian S, Lord EM. 2010. A multifaceted study of stigma/style cysteine-rich adhesin (SCA)-like Arabidopsis lipid transfer proteins (LTPs) suggests diversified roles for these LTPs in plant growth and reproduction. *Journal of Experimental Botany* **61**(15): 4277-4290.
Chae K, Kieslich CA, Morikis D, Kim SC, Lord EM. 2009. A gain-of-function mutation of Arabidopsis lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. *Plant Cell* **21**(12): 3902-3914.

Chapman LA, Goring DR. 2010. Pollen-pistil interactions regulating successful fertilization in the Brassicaceae. *Journal of Experimental Botany* **61**(7): 1987-1999.

Charlesworth D, Bartolome C, Schierup MH, Mable BK. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Molecular Biology and Evolution* **20**(11): 1741-1753.

Charlesworth, D, Guttman, DS (1999). The evolution of dioecy and plant sex chromosome systems. In: Ainsworth CC (ed), *Sex Determination in Plants. BIOS, Oxford,* pp 25–49.

Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**: 237-268.

Chen KY, Cong B, Wing R, Vrebalov J, Tanksley SD. 2007. Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science* **318**(5850): 643-645.

Chen C, Chen G, Hao X, Cao B, Chen Q, Liu S, Lei J. 2011. *CaMF2*, an anther-specific lipid transfer protein (LTP) gene, affects pollen development in *Capsicum annuum* L. *Plant Science* **181**(4): 439-448.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: A program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology* **7**(3-4): 429-447.

Chen ZX. 2001. A superfamily of proteins with novel cysteine-rich repeats. *Plant Physiology* **126**(2): 473-476.

Chookajorn T, Kachroo A, Ripoll DR, Clark AG, Nasrallah JB. 2004. Specificity determinants and diversification of the *Brassica* self-incompatibility pollen ligand. *Proceedings of the National Academy of Sciences of the United States of America* **101**(4): 911-917.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* **131**(1): 11-22.

Clark SE, Running MP, Meyerowitz EM. 1993. CLAVATA1, a regulator of meristem and flower development in *Arabidopsis*. *Development* **119**(2): 397-418.

Clark SE, Running MP, Meyerowitz EM. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* **121**(7): 2057-2067.

Clarke A, Gleeson P, Harrison S, Knox RB. 1979. Pollen-stigma interactions - identification and characterization of surface components with recognition potential. *Proceedings of the National Academy of Sciences of the United States of America* **76**(7): 3358-3362.

Combier J-P, Kuster H, Journet E-P, Hohnjec N, Gamas P, Niebel A. 2008. Evidence for the involvement in nodulation of the two small putative regulatory peptide-encoding genes *MtRALFL1* and *MtDVL1*. *Molecular plant-microbe interactions : MPMI* **21**(8): 1118-1127.

Costa LM, Marshall E, Tesfaye M, Silverstein KAT, Mori M, Umetsu Y, Otterbach SL, Papareddy R, Dickinson HG, Boutiller K, *et al.* 2014. Central cell-derived peptides regulate early embryo patterning in flowering plants. *Science* **344**(6180): 168-172.

Covey PA, Kondo K, Welch L, Frank E, Sianta S, Kumar A, Nunez R, Lopez-Casado G, van der Knaap E, Rose JKC, *et al*. 2010. Multiple features that distinguish unilateral incongruity and self-incompatibility in the tomato clade. *Plant Journal* **64**(3): 367-378.

Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, Zhang D. 2015. Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Molecular Plant* **8**(6): 935-945.

Dai S, Li L, Chen T, Chong K, Xue Y, Wang T. 2006. Proteomic analyses of *Oryza sativa* mature pollen reveal novel proteins associated with pollen germination and tube growth. *Proteomics* **6**(8): 2504-2529.

Darwin F, Seward AC. 1903. More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters. *More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters*.

de Beer A, Vivier MA. 2011. Four plant defensins from an indigenous South African Brassicaceae species display divergent activities against two test pathogens despite high sequence similarity in the encoding genes. *BMC research notes* **4**: 459-459.

de Silva K, Laska B, Brown C, Sederoff HW, Khodakovskaya M. 2011. *Arabidopsis thaliana* calcium-dependent lipid-binding protein (AtCLB): a novel repressor of abiotic stress response. *Journal of Experimental Botany* **62**(8): 2679-2689.

Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* **64**(7): 2044-2054.

Dickinson H, Lewis D. 1973. Formation of tryphine coating pollen grains of *Raphanus*, and its properties relating to self-incompatibility system. *Proceedings of the Royal Society Series B-Biological Sciences* **184**(1075): 149-165.

Dickinson H. 1995. Dry Stigmas, Water and Self-Incompatibility in *Brassica*. *Sexual Plant Reproduction* **8**(1): 1-10.

Dobritsa AA, Nishikawa SI, Preuss D, Urbanczyk-Wochniak E, Sumner LW, Hammond A, Carlson AL, Swanson RJ. 2009. LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. *Sexual Plant Reproduction* **22**(3): 167-177.

Doughty J, Dixon S, Hiscock SJ, Willis AC, Parkin IAP, Dickinson HG. 1998. PCP-A1, a defensin-like Brassica pollen coat protein that binds the *S* locus glycoprotein, is the product of gametophytic gene expression. *Plant Cell* **10**(8): 1333-1347.

Doughty J, Hedderson F, Mccubbin A, Dickinson H. 1993. Interaction between a coating-borne peptide of the Brassica pollen grain and stigmatic-*S* (self-incompatibility)-locus-specific glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **90**(2): 467-471.

Doughty J, Wong HY, Dickinson HG. 2000. Cysteine-rich pollen coat proteins (PCPs) and their interactions with stigmatic *S* (incompatibility) and *S*-related proteins in *Brassica*: putative roles in SI and pollination. *Annals of Botany* **85**: 161-169.

Dresselhaus T, Franklin-Tong N. 2013. Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization. *Molecular Plant* **6**(4): 1018-1036.

Dresselhaus T, Lorz H, Kranz E. 1994. Representative cDNA libraries from few plant-cells. *Plant Journal* **5**(4): 605-610.

Eaves DJ, Flores-Ortiz C, Haque T, Lin ZC, Teng NJ, Franklin-Tong VE. 2014. Self-incompatibility in *Papaver*: advances in integrating the signalling network. *Biochemical Society Transactions* **42**: 370-376.

Edlund AF, Swanson R, Preuss D. 2004. Pollen and stigma structure and function: the role of diversity in pollination. *Plant Cell* **16**: S84-S97.

Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA-sequences in eukaryotes. *Quarterly Review of Biology* **70**(3): 297-320.

Elleman CJ, Dickinson HG. 1986. Pollen-stigma interactions in *Brassica*. IV. Structural reorganization in the pollen grains during hydration. *Journal of Cell Sciences* **80**: 141-157.

Elleman CJ, Dickinson HG. 1996. Identification of pollen components regulating pollination-specific responses in the stigmatic papillae of *Brassica oleracea*. *New Phytologist* **133**(2): 197-205.

Elleman CJ, Dickinson HG. 1999. Commonalties between pollen/stigma and host/pathogen interactions: calcium accumulation during stigmatic penetration by *Brassica oleracea* pollen tubes. *Sexual Plant Reproduction* **12**(3): 194-202.

Elleman CJ, Franklin-Tong V, Dickinson HG. 1992. Pollination in species with dry stigmas - the nature of the early stigmatic response and the pathway taken by pollen tubes. *New Phytologist* **121**(3): 413-424.

Endo S, Betsuyaku S, Fukuda H. 2014. Endogenous peptide ligand-receptor systems for diverse signaling networks in plants. *Current Opinion in Plant Biology* **21**: 140-146.

Entani T, Iwano M, Shiba H, Che FS, Isogai A, Takayama S. 2003. Comparative analysis of the self-incompatibility (*S*-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes to Cells* **8**(3): 203-213.

Evans DE, Taylor PE, Singh MB, Knox RB. 1992. The interrelationship between the accumulation of lipids, protein and the level of acyl carrier protein during the development of *Brassica napus* l pollen. *Planta* **186**(3): 343-354.

Fant F, Vranken W, Broekaert W, Borremans F. 1998. Determination of the three-dimensional solution structure of *Raphanus sativus* antifungal protein 1 by H-1 NMR. *Journal of Molecular Biology* **279**(1): 257-270.

Fant F, Vranken WF, Borremans FAM. 1999. The three-dimensional solution structure of *Aesculus hippocastanum* antimicrobial protein 1 determined by H-1 nuclear magnetic resonance. *Proteins-Structure Function and Genetics* **37**(3): 388-403.

Farkas A, Maroti G, Durgo H, Gyorgypal Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert P, Kondorosi E. 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **111**(14): 5183-5188.

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP. 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiology* **130**(2): 519-537.

Fiebig A, Mayfield JA, Miley NL, Chau S, Fischer RL, Preuss D. 2000. Alterations in *CER6*, a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* **12**(10): 2001-2008.

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong W-K, Mockler TC. 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Research* **20**(1): 45-58.

Fisher RA. 1941. Average excess and average effect of a gene substitution. *Annals of Human Genetics* **11**(1): 53-63.

Foote HCC, Ride JP, Franklintong VE, Walker EA, Lawrence MJ, Franklin FCH. 1994. Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L. *Proceedings of the National Academy of Sciences of the United States of America* **91**(6): 2265-2269.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**(4): 1531-1545.

Foster GD, Robinson SW, Blundell RP, Roberts MR, Hodge R, Draper J, Scott RJ. 1992. A *Brassica napus* messenger-RNA encoding a protein homologous to phospholipid transfer proteins, is expressed specifically in the tapetum and developing microspores. *Plant Science* **84**(2): 187-192.

Franklin-Tong VE, Holdaway-Clarke TL, Straatman KR, Kunkel JG, Hepler PK. 2002. Involvement of extracellular calcium influx in the self-incompatibility response of *Papaver rhoeas*. *Plant Journal* **29**(3): 333-345.

Friedman WE. 1998. The evolution of double fertilization and endosperm: an "historical" perspective. *Sexual Plant Reproduction* **11**(1): 6-16.

Friis EM, Pedersen KR, Crane PR. 2006. Cretaceous anglosperin flowers: innovation and evolution in plant reproduction. *Palaeogeography Palaeoclimatology Palaeoecology* **232**(2-4): 251-293.

Fujimoto R, Sugimura T, Nishio T. 2006. Gene conversion from *SLG* to *SRK* resulting in self-compatibility in *Brassica rapa*. *FEBS Letters* **580**(2): 425-430.

Galindo BE, Moy GW, Swanson WJ, Vacquier VD. 2002. Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. *Gene* **288**(1-2): 111-117.

Galindo BE, Vacquier VD, Swanson WJ. 2003. Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences of the United States of America* **100**(8): 4639-4643.

Gaude T, Dumas C. 1984. A membrane-like structure on the pollen wall surface in *Brassica*. *Annals of Botany* **54**(6): 821-825.

Ge LL, Tian HQ, Russell SD. 2007. Calcium function and distribution during fertilization in angiosperms. *American Journal of Botany* **94**(6): 1046-1060.

Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, Choulet F. 2015. Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biology* **16**:188.

Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B. 2010. Species selection maintains self-incompatibility. *Science* **330**(6003): 493-495.

Goldman N, Yang ZH. 1994. Codon-based model of nucleotide substitution for protein-coding DNA-sequences. *Molecular Biology and Evolution* **11**(5): 725-736.

Goodwillie C, Kalisz S, Eckert CG. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology Evolution and Systematics* **36**: 47-79.

Graham MA, Silverstein KAT, Cannon SB, VandenBosch KA. 2004. Computational identification and characterization of novel genes from legumes. *Plant Physiology* **135**(3): 1179-1197.

Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30**: S162-S173.

Guo Y-L, Zhao X, Lanz C, Weigel D. 2011. Evolution of the *S*-locus region in *Arabidopsis* relatives. *Plant Physiology* **157**(2): 937-946.

Gutierrez-Marcos JF, Costa LM, Biderre-Petit C, Khbaya B, O'Sullivan DM, Wormald M, Perez P, Dickinson HG. 2004. Maternally expressed gene1 is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell* **16**(5): 1288-1301.

Haasen KE, Goring DR. 2010. The recognition and rejection of self-incompatible pollen in the Brassicaceae. *Botanical Studies* **51**(1): 1-6.

Habeeb AFS, Hiramoto R. 1968. Reaction of proteins with glutaraldehyde. *Archives of Biochemistry and Biophysics* **126**(1): 16-26.

Hammami R, Ben Hamida J, Vergoten G, Fliss I. 2009. PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Research* **37**: D963-D968.

Hara K, Yokoo T, Kajita R, Onishi T, Yahata S, Peterson KM, Torii KU, Kakimoto T. 2009. Epidermal cell density is autoregulated via a secretory peptide, epidermal patterning factor 2 in Arabidopsis leaves. *Plant and Cell Physiology* **50**(6): 1019-1031.

Haruta M, Sabat G, Stecker K, Minkoff BB, Sussman MR. 2014. A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science* **343**(6169): 408-411.

He B, Guo W. 2009. The exocyst complex in polarized exocytosis. *Current Opinion in Cell Biology* **21**(4): 537-542.

Hernandez-Pinzon I, Ross JHE, Barnes KA, Damant AP, Murphy DJ. 1999. Composition and role of tapetal lipid bodies in the biogenesis of the pollen coat of *Brassica napus*. *Planta* **208**(4): 588-598.

Heslop-Harrison J. 1968. Tapetal origin of pollen-coat substances in *Lilium*. *New Phytologist* **67**(4): 779-786.

Heslop-Harrison Y, Shivanna KR. 1977. Receptive surface of angiosperm stigma. *Annals of Botany* **41**(176): 1233-1258.

Higashiyama T, Takeuchi H. 2015. The mechanism and key molecules involved in pollen tube guidance. *Annual Review of Plant Biology, Vol 66* **66**: 393-413.

Higashiyama T, Yabe S, Sasaki N, Nishimura Y, Miyagishima S, Kuroiwa H, Kuroiwa T. 2001. Pollen tube attraction by the synergid cell. *Science* **293**(5534): 1480-1483.

Hiscock SJ, Allen AM. 2008. Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus. *New Phytologist* **179**(2): 286-317.

Hiscock SJ, Dickinson HG. 1993. Unilateral incompatibility within the Brassicaceae - further evidence for the involvement of the self-incompatibility (*S*)-locus. *Theoretical and Applied Genetics* **86**(6): 744-753.

Hiscock SJ, Doughty J, Willis AC, Dickinson HG. 1995. A 7-kDa pollen coating-borne peptide from *Brassica napus* interacts with *S*-locus glycoprotein and *S*-locus-related glycoprotein. *Planta* **196**(2): 367-374.

Hiscock SJ, Hoedemaekers K, Friedman WE, Dickinson HG. 2002. The stigma surface and pollen-stigma interactions in *Senecio squalidus* L. (Asteraceae) following cross (compatible) and self (incompatible) pollinations. *International Journal of Plant Sciences* **163**(1): 1-16.

Hiscock SJ, Dewey FM, Coleman JOD, Dickinson HG. 1994. Identification and localization of an active cutinase in the pollen of *Brassica napus* L. *Planta* **193**(3): 377-384.

Hiscock SJ, Bown D, Gurr SJ, Dickinson HG. 2002. Serine esterases are required for pollen tube penetration of the stigma in *Brassica*. *Sexual Plant Reproduction* **15**(2): 65-74.

Hochuli PA, Feist-Burkhardt S. 2013. Angiosperm-like pollen and Afropollis from the Middle Triassic (Anisian) of the Germanic Basin (Northern Switzerland). *Frontiers in Plant Science* **4**: 344.

Huang M-D, Chen T-LL, Huang AHC. 2013. Abundant type III lipid transfer proteins in Arabidopsis tapetum are secreted to the locule and become a constituent of the pollen exine. *Plant Physiology* **163**(3): 1218-1229.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society B-Biological Sciences* **256**(1346): 119-124.

Hulskamp M, Kopczak SD, Horejsi TF, Kihl BK, Pruitt RE. 1995. Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. *Plant Journal* **8**(5): 703-714.

Husband BC, Sabara HA. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* **161**(3): 703-713.

Ikeda K, Igic B, Ushijima K, Yamane H, Hauck NR, Nakano R, Sassa H, Iezzoni AF, Kohn JR, Tao R. 2004. Primary structural features of the *S* haplotype-specific F-box protein, SFB, in *Prunus*. *Sexual Plant Reproduction* **16**(5): 235-243.

Indriolo E, Goring DR. 2014. A conserved role for the ARC1 E3 ligase in Brassicaceae self-incompatibility. *Frontiers in Plant Science* **5**: 181.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* **11**(2): 97-108.

Iwano M, Ito K, Fujii S, Kakita M, Asano-Shimosato H, Igarashi M, Kaothien-Nakayama P, Entani T, Kanatani A, Takehisa M, *et al*. 2015. Calcium signalling mediates self-incompatibility response in the Brassicaceae. *Nature Plants* **1**(9): 15128.

Iwano M, Igarashi M, Tarutani Y, Kaothien-Nakayama P, Nakayama H, Moriyama H, Yakabe R, Entani T, Shimosato-Asano H, Ueki M, *et al*. 2014. A Pollen Coat-inducible autoinhibited Ca^{2+}-ATPase expressed in stigmatic papilla cells is required for compatible pollination in the Brassicaceae. *Plant Cell* **26**(2): 636-649.

Iwano M, Shiba H, Matoba K, Miwa T, Funato M, Entani T, Nakayama P, Shimosato H, Takaoka A, Isogai A, *et al*. 2007. Actin dynamics in papilla cells of *Brassica rapa* during self- and cross-pollination. *Plant Physiology* **144**(1): 72-81.

Iwano M, Shiba H, Miwa T, Che FS, Takayama S, Nagai T, Miyawaki A, Isogai A. 2004. Ca^{2+} dynamics in a pollen grain and papilla cell during pollination of *Arabidopsis*. *Plant Physiology* **136**(3): 3562-3571.

Iwano M, Takayama S. 2012. Self/non-self discrimination in angiosperm self-incompatibility. *Current Opinion in Plant Biology* **15**(1): 78-83.

Iwano M, Wada M, Morita Y, Shiba H, Takayama S, Isogai A. 1999. X-ray microanalysis of papillar cells and pollen grains in the pollination process in *Brassica* using a variable-pressure scanning electron microscope. *Journal of Electron Microscopy* **48**(6): 909-917.

Jang CS, Yim WC, Moon J-C, Jung JH, Lee TG, Lim SD, Cho SH, Lee KK, Kim W, Seo YW, *et al*. 2008. Evolution of non-specific lipid transfer protein (nsLTP) genes in the Poaceae family: their duplication and diversity. *Molecular Genetics and Genomics* **279**(5): 481-497.

Jauh GY, Eckard KJ, Nothnagel EA, Lord EM. 1997. Adhesion of lily pollen tubes on an artificial matrix. *Sexual Plant Reproduction* **10**(3): 173-180.

Jewaria PK, Hara T, Tanaka H, Kondo T, Betsuyaku S, Sawa S, Sakagami Y, Aimoto S, Kakimoto T. 2013. Differential effects of the peptides stomagen, EPF1 and EPF2 on activation of MAP Kinase MPK6 and the SPCH protein level. *Plant and Cell Physiology* **54**(8): 1253-1262.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, *et al*. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**(7345): 97-100.

Jones-Rhoades MW, Borevitz JO, Preuss D. 2007. Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *Plos Genetics* **3**(10): 1848-1861.

Joshi AK, Chand R, Kumar S, Singh RP. 2004. Leaf tip necrosis: a phenotypic marker associated with resistance to spot blotch disease in wheat. *Crop Science* **44**(3): 792-796.

Jung HW, Lim CW, Hwang BK. 2006. Isolation and functional analysis of a pepper lipid transfer protein III (CALTPIII) gene promoter during signaling to pathogen, abiotic and environmental stresses. *Plant Science* **170**(2): 258-266.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics* **10**(1): 19-31.

Kakita M, Murase K, Iwano M, Matsumoto T, Watanabe M, Shiba H, Isogai A, Takayama S. 2007. Two distinct forms of *M*-locus protein kinase localize to the plasma membrane and interact directly with *S*-locus receptor kinase to transduce self-incompatibility signaling in *Brassica rapa*. *Plant Cell* **19**(12): 3961-3973.

Kao TH, Tsukamoto T. 2004. The molecular and genetic bases of *S*-RNase-based self-incompatibility. *Plant Cell* **16**: S72-S83.
Kato S, Mukai Y. 2004. Allelic diversity of *S*-RNase at the self-incompatibility locus in natural flowering cherry populations (*Prunus lannesiana* var. *speciosa*). *Heredity* **92**(3): 249-256.
Kay KM. 2006. Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. *Evolution* **60**(3): 538-552.
Kemp BP, Doughty J. 2007. *S* cysteine-rich (SCR) binding domain analysis of the Brassica self-incompatibility S-locus receptor kinase. *New Phytologist* **175**(4): 619-629.
Kevei Z, Vinardell JM, Kiss GB, Kondorosi A, Kondorosi E. 2002. Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in *Medicago* spp. *Molecular Plant-Microbe Interactions* **15**(9): 922-931.
Kiefer F, Arnold K, Kuenzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research* **37**: D387-D392.
Kim JY, Park SC, Hwang I, Cheong H, Nah JW, Hahm KS, Park Y. 2009. Protease inhibitors from plants with antimicrobial activity. *International Journal of Molecular Sciences* **10**(6): 2860-2872.
Kitashiba H, Nasrallah JB. 2014. Self-incompatibility in Brassicaceae crops: lessons for interspecific incompatibility. *Breeding Science* **64**(1): 23-37.
Kleinboelting N, Huep G, Kloetgen A, Viehoever P, Weisshaar B. 2012. GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Research* **40**(D1): D1211-D1215.
Knox RB, Heslop-Harrison J. 1970. Pollen-wall proteins: localization and enzymic activity. *Journal of Cell Science* **6**(1): 1-27.
Kondo T, Kajita R, Miyazaki A, Hokoyama M, Nakamura-Miura T, Mizuno S, Masuda Y, Irie K, Tanaka Y, Takada S, *et al*. 2010. Stomatal density is controlled by a mesophyll-derived signaling molecule. *Plant and Cell Physiology* **51**(1): 1-8.
Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. *Journal of Theoretical Biology* **239**(2): 141-151.
Kondrashov FA, Koonin EV. 2004. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends in Genetics* **20**(7): 287-291.
Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biology* **3**(2): RESEARCH0008.
Lacerda AF, Vasconcelos EAR, Pelegrini PB, Grossi de Sa MF. 2014. Antifungal defensins and their role inplant defense. *Frontiers in Microbiology* **5**: 116.
Lai Z, Ma WS, Han B, Liang LZ, Zhang YS, Hong GF, Xue YB. 2002. An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Molecular Biology* **50**(1): 29-42.
Lampard GR, MacAlister CA, Bergmann DC. 2008. Arabidopsis stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. *Science* **322**(5904): 1113-1116.
Lavithis M, Bhalla PL. 1995. Esterases in pollen and stigma of *Brassica*. *Sexual Plant Reproduction* **8**(5): 289-298.
Le Q, Gutierrez-Marcos JF, Costa LM, Meyer S, Dickinson HG, Lorz H, Kranz E, Scholten S. 2005. Construction and screening of subtracted cDNA libraries from limited populations of plant cells: a comparative analysis of gene expression between maize egg cells and central cells. *Plant Journal* **44**(1): 167-178.
Lease KA, Walker JC. 2006. The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiology* **142**(3): 831-838

Lecourieux D, Raneva R, Pugin A. 2006. Calcium in plant defence-signalling pathways. *New Phytologist* **171**(2): 249-269.
Lessios HA. 2011. Speciation genes in free-spawning marine invertebrates. *Integrative and Comparative Biology* **51**(3): 456-465.
Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**(1): 127-128.
Lewis CA, Talbot CF, Vacquier VD. 1982. A protein from abalone sperm dissolves the egg vitelline layer by a non-enzymatic mechanism. *Developmental Biology* **92**(1): 227-239.
Li WH, Gojobori T, Nei M. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature* **292**(5820): 237-239.
Li WT, Chetelat RT. 2014. The role of a pollen-expressed cullin1 protein in gametophytic self-incompatibility in *Solanum*. *Genetics* **196**(2): 439-442.
Li WT, Chetelat RT. 2015. Unilateral incompatibility gene UI1.1 encodes an *S*-locus F-box protein expressed in pollen of *Solanum* species. *Proceedings of the National Academy of Sciences of the United States of America* **112**(14): 4417-4422.
Liu F, Zhang X, Lu C, Zeng X, Li Y, Fu D, Wu G. 2015. Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *Journal of Experimental Botany* **66**(19): 5663-5681.
Liu JJ, Zhong S, Guo XY, Hao LH, Wei XL, Huang QP, Hou YN, Shi J, Wang CY, Gu HY, *et al*. 2013. Membrane-bound RLCKs LIP1 and LIP2 are essential male factors controlling male-female attraction in *Arabidopsis*. *Current Biology* **23**(11): 993-998.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**(5494): 1151-1155.
Magnard JL, Le Deunff E, Domenech J, Rogowsky PM, Testillano PS, Rougier M, Risueno MC, Vergne P, Dumas C. 2000. Genes normally expressed in the endosperm are expressed at early stages of microspore embryogenesis in maize. *Plant Molecular Biology* **44**(4): 559-574.
Mano S, Innan H. 2008. The evolutionary rate of duplicated genes under concerted evolution. *Genetics* **180**(1): 493-505.
Marshall E, Costa LM, Gutierrez-Marcos J. 2011. Cysteine-Rich Peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *Journal of Experimental Botany* **62**(5): 1677-1686.
Martin NH, Willis JH. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* **61**(1): 68-82.
Maruyama D, Sugiyama T, Endo T, Nishikawa S. 2014. Multiple *BiP* Genes of *Arabidopsis thaliana* are required for male gametogenesis and pollen competitiveness. *Plant and Cell Physiology* **55**(4): 801-810.
Matsubayashi Y. 2003. Ligand-receptor pairs in plant peptide signaling. *Journal of Cell Science* **116**(19): 3863-3870.
Mayfield JA, Fiebig A, Johnstone SE, Preuss D. 2001. Gene families from the *Arabidopsis thaliana* pollen coat proteome. *Science* **292**(5526): 2482-2485.
Mayfield JA, Preuss D. 2000. Rapid initiation of Arabidopsis pollination requires the oleosin-domain protein GRP17. *Nature Cell Biology* **2**(2): 128-130.
McClure B. 2004. *S*-RNase and SLF determine *S*-haplotype-specific pollen recognition and rejection. *Plant Cell* **16**(11): 2840-2847.
McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE. 1989. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature* **342**(6252): 955-957.

McInnis SM, Desikan R, Hancock JT, Hiscock SJ. 2006. Production of reactive oxygen species and reactive nitrogen species by angiosperm stigmas and pollen: potential signalling crosstalk? *New Phytologist* **172**(2): 221-228.

Meindl T, Boller T, Felix G. 1998. The plant wound hormone systemin binds with the N-terminal part to its receptor but needs the C-terminal part to activate it. *Plant Cell* **10**(9): 1561-1570.

Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E. 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiology* **132**(1): 161-173.

Metz EC, Palumbi SR. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Molecular Biology and Evolution* **13**(2): 397-406.

Metz EC, Robles-Sikisaka R, Vacquier VD. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **95**(18): 10676-10681.

Mi HY, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols* **8**(8): 1551-1566.

Mishler BD, Donoghue MJ. 1982. Species concepts - a case for pluralism. *Systematic Zoology* **31**(4): 491-503.

Mollet JC, Park SY, Nothnagel EA, Lord EM. 2000. A lily stylar pectin is necessary for pollen tube adhesion to an in vitro stylar matrix. *Plant Cell* **12**(9): 1737-1749.

Murphy DJ. 2006. The extracellular pollen coat in members of the Brassicaceae: composition, biosynthesis, and functions in pollination. *Protoplasma* **228**(1-3): 31-39.

Murphy DJ, Ross JHE. 1998. Biosynthesis, targeting and processing of oleosin-like proteins, which are major pollen coat components in *Brassica napus*. *Plant Journal* **13**(1): 1-16.

Murphy E, De Smet I. 2014. Understanding the RALF family: a tale of many species. *Trends in Plant Science* **19**(10): 664-671.

Muschietti J, Dircks L, Vancanneyt G, McCormick S. 1994. Lat52 protein is essential for tomato pollen development - pollen expressing antisense *Lat52* RNA hydrates and germinates abnormally and cannot achieve fertilization. *Plant Journal* **6**(3): 321-338.

Nalefski EA, Falke JJ. 1996. The C2 domain calcium-binding motif: structural and functional diversity. *Protein Science* **5**(12): 2375-2390.

Nasrallah JB. 1997. Evolution of the *Brassica* self-incompatibility locus: a look into *S*-locus gene polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **94**(18): 9516-9519.

Nielsen R, Yang ZH. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**(3): 929-936.

Nishikawa SI, Zinkl GM, Swanson RJ, Maruyama D, Preuss D. 2005. Callose (β-1,3 glucan) is essential for *Arabidopsis* pollen wall patterning, but not tube growth. *BMC Plant Biology* **5**: 22.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**(1): 205-217.

Ohno S. 1970. Evolution by gene duplication. Berlin, Germany: *Springer*: 1-160.

Oikawa E, Takuno S, Izumita A, Sakamoto K, Hanzawa H, Kitashiba H, Nishio T. 2011. Simple and efficient methods for *S* genotyping and *S* screening in genus *Brassica* by dot-blot analysis. *Molecular Breeding* **28**(1): 1-12.

Okuda S, Tsutsui H, Shiina K, Sprunck S, Takeuchi H, Yui R, Kasahara RD, Hamamura Y, Mizukami A, Susaki D, *et al*. 2009. Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**(7236): 357-361.

Onus AN, Pickersgill B. 2004. Unilateral incompatibility in *Capsicum* (Solanaceae): occurrence and taxonomic distribution. *Annals of Botany* **94**(2): 289-295.

Osborn RW, Desamblanx GW, Thevissen K, Goderis I, Torrekens S, Vanleuven F, Attenborough S, Rees SB, Broekaert WF. 1995. Isolation and characterization of plant defensins from seeds of Asteraceae, Fabaceae, Hippocastanaceae and Saxifragaceae. *FEBS Letters* **368**(2): 257-262.

Park CJ, Seo YS. 2015. Heat shock proteins: a review of the molecular chaperones for plant immunity. *Plant Pathology Journal* **31**(4): 323-333.

Park SY, Jauh GY, Mollet JC, Eckard KJ, Nothnagel EA, Walling LL, Lord EM. 2000. A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix. *Plant Cell* **12**(1): 151-163.

Park SY, Lord EM. 2003. Expression studies of *SCA* in lily and confirmation of its role in pollen tube adhesion. *Plant Molecular Biology* **51**(2): 183-189.

Paul W, Hodge R, Smartt S, Draper J, Scott R. 1992. The isolation and characterization of the tapetum-specific *Arabidopsis thaliana A9* gene. *Plant Molecular Biology* **19**(4): 611-622.

Pearce G, Moura DS, Stratmann J, Ryan CA. 2001. RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. *Proceedings of the National Academy of Sciences of the United States of America* **98**(22): 12843-12847.

Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. 2007. Genes under positive selection in *Escherichia coli*. *Genome Research* **17**(9): 1336-1343.

Petrov V, Hille J, Mueller-Roeber B, Gechev TS. 2015. ROS-mediated abiotic stress-induced programmed cell death in plants. *Frontiers in Plant Science* **6**:69.

Piffanelli P, Ross JHE, Murphy DJ. 1998. Biogenesis and function of the lipidic structures of pollen grains. *Sexual Plant Reproduction* **11**(2): 65-80.

Pillitteri LJ, Torii KU. 2012. Mechanisms of stomatal development. *Annual Review of Plant Biology* **63**: 591-614.

Preuss D, Lemieux B, Yen G, Davis RW. 1993. A conditional sterile mutation eliminates surface components from Arabidopsis pollen and disrupts cell signaling during fertilization. *Genes and Development* **7**(6): 974-985.

Punwani JA, Rabiger DS, Drews GN. 2007. MYB98 positively regulates a battery of synergid-expressed genes encoding filiform apparatus-localized proteins. *Plant Cell* **19**(8): 2557-2568.

Punwani JA, Rabiger DS, Lloyd A, Drews GN. 2008. The *MYB98* subcircuit of the synergid gene regulatory network includes genes directly and indirectly regulated by MYB98. *Plant Journal* **55**(3): 406-414.

Qi YP, Katagiri F. 2009. Purification of low-abundance Arabidopsis plasma-membrane protein complexes and identification of candidate components. *Plant Journal* **57**(5): 932-944.

Quilichini TD, Douglas CJ, Samuels AL. 2014. New views of tapetum ultrastructure and pollen exine development in *Arabidopsis thaliana*. *Annals of Botany* **114**(6): 1189-1201.

Rademacher S, Sprunck S. 2013. Downregulation of egg cell-secreted EC1 is accompanied with delayed gamete fusion and polytubey. *Plant signaling & behavior* **8**(12): e27377-e27377.

Ramsey J, Bradshaw HD, Schemske DW. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**(7): 1520-1534.

Renner SS, Ricklefs RE. 1995. Dioecy and its correlates in the flowering plants. *American Journal of Botany* **82**(5): 596-606.

Rieseberg LH, Blackman BK. 2010. Speciation genes in plants. *Annals of Botany* **106**(3): 439-455.

Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* **317**(5840): 910-914.

Rieseberg LH, Wood TE, Baack EJ. 2006. The nature of plant species. *Nature* **440**(7083): 524-527.

Rodriguez L, Gonzalez-Guzman M, Diaz M, Rodrigues A, Izquierdo-Garcia AC, Peirats-Llobet M, Fernandez MA, Antoni R, Fernandez D, Marquez JA, *et al.* 2014. C2-domain abscisic acid-related proteins mediate the interaction of Pyr/Pyl/Rcar abscisic acid receptors with the plasma membrane and regulate abscisic acid sensitivity in *Arabidopsis*. *Plant Cell* **26**(12): 4802-4820.

Ross JHE, Murphy DJ. 1996. Characterization of anther-expressed genes encoding a major class of extracellular oleosin-like proteins in the pollen coat of Brassicaceae. *Plant Journal* **9**(5): 625-637.

Safavian D, Goring DR. 2013. Secretory activity is rapidly induced in stigmatic papillae by compatible pollen, but inhibited for self-incompatible pollen in the Brassicaceae. *Plos One* **8**(12): e84286.

Sagaram US, El-Mounadi K, Buchko GW, Berg HR, Kaur J, Pandurangi RS, Smith TJ, Shah DM. 2013. Structural and functional studies of a phosphatidic acid-binding antifungal plant defensin MtDef4: Identification of an RGFRRR motif governing fungal cell entry. *Plos One* **8**(12):e82485.

Samuel MA, Chong YT, Haasen KE, Aldea-Brydges MG, Stone SL, Goring DR. 2009. Cellular pathways regulating responses to compatible and self-incompatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. *Plant Cell* **21**(9): 2655-2671.

Samuel MA, Tang W, Jamshed M, Northey J, Patel D, Smith D, Siu KWM, Muench DG, Wang Z-Y, Goring DR. 2011. Proteomic analysis of *Brassica* stigmatic proteins following the self-incompatibility reaction reveals a role for microtubule dynamics during pollen responses. *Molecular & Cellular Proteomics* **10**(12): M111.00138.

Sancho AI, Rigby NM, Zuidmeer L, Asero R, Mistrello G, Amato S, Gonzalez-Mancebo E, Fernandez-Rivas M, Ree R, Mills ENC. 2005. The effect of thermal processing on the IgE reactivity of the non-specific lipid transfer protein from apple, Mal d 3. *Allergy* **60**(10): 1262-1268.

Sato K, Nishio T, Kimura R, Kusaba M, Suzuki T, Hatakeyama K, Ockendon DJ, Satta Y. 2002. Coevolution of the *S*-locus genes *SRK*, *SLG* and *SP11/SCR* in *Brassica oleracea* and *B. rapa*. *Genetics* **162**(2): 931-940.

Scheer JM, Ryan CA. 2002. The systemin receptor SR160 from *Lycopersicon peruvianum* is a member of the LRR receptor kinase family. *Proceedings of the National Academy of Sciences of the United States of America* **99**(14): 9585-9590.

Scheres B, Vanengelen F, Vanderknaap E, Vandewiel C, Vankammen A, Bisseling T. 1990. Sequential induction of nodulin gene-expression in the developing pea nodule. *Plant Cell* **2**(8): 687-700.

Schopfer CR, Nasrallah ME, Nasrallah JB. 1999. The male determinant of self-incompatibility in *Brassica*. *Science* **286**(5445): 1697-1700.

Schwartz C, Balasubramanian S, Warthmann N, Michael TP, Lempe J, Sureshkumar S, Kobayashi Y, Maloof JN, Borevitz JO, Chory J, *et al*. 2009. *Cis*-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics* **183**(2): 723-732.

Serna A, Maitz M, O'Connell T, Santandrea G, Thevissen K, Tienens K, Hueros G, Faleri C, Cai G, Lottspeich F, *et al*. 2001. Maize endosperm secretes a novel antifungal protein into adjacent maternal tissue. *Plant Journal* **25**(6): 687-698.

Shiba H, Takayama S, Iwano M, Shimosato H, Funato M, Nakagawa T, Che FS, Suzuki G, Watanabe M, Hinata K, *et al*. 2001. A pollen coat protein, SP11/SCR, determines the pollen *S*-specificity in the self-incompatibility of *Brassica* species. *Plant Physiology* **125**(4): 2095-2103.

Shiu, S. H., and Bleecker, A. B. 2001. Plant receptor-like kinase gene family: diversity, function, and signaling. *Science's STKE* 2001:re22.

Shpak ED, McAbee JM, Pillitteri LJ, Torii KU. 2005. Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science* **309**(5732): 290-293.

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao TH. 2004. Identification of the pollen determinant of *S*-RNase-mediated self-incompatibility. *Nature* **429**(6989): 302-305.

Silverstein KAT, Graham MA, Paape TD, VandenBosch KA. 2005. Genome organization of more than 300 defensin-like genes in Arabidopsis. *Plant Physiology* **138**(2): 600-610.

Silverstein KAT, Moskal WA, Jr., Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA. 2007. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant Journal* **51**(2): 262-280.

Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **99**(21): 13627-13632.

Smyth DR, Bowman JL, Meyerowitz EM. 1990. Early flower development in Arabidopsis. *Plant Cell* **2**(8): 755-767.

Sprunck S, Baumann U, Edwards K, Langridge P, Dresselhaus T. 2005. The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant Journal* **41**(5): 660-672.

Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U, Dresselhaus T. 2012. Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* **338**(6110): 1093-1097.

Srivastava R, Liu J-X, Guo H, Yin Y, Howell SH. 2009. Regulation and processing of a plant peptide hormone, AtRALF23, in Arabidopsis. *Plant Journal* **59**(6): 930-939.

Stael S, Kmiecik P, Willems P, Van Der Kelen K, Coll NS, Teige M, Van Breusegem F. 2015. Plant innate immunity - sunny side up? *Trends in Plant Science* **20**(1): 3-11.

Stebbins GL. 1957. Self fertilization and population variability in the higher plants. *American Naturalist* **91**(861): 337-354.

Steffen JG, Kang I-H, Macfarlane J, Drews GN. 2007. Identification of genes expressed in the Arabidopsis female gametophyte. *Plant Journal* **51**(2): 281-292.

Steinhorst L, Kudla J. 2013. Calcium - a central regulator of pollen germination and tube growth. *Biochimica Et Biophysica Acta-Molecular Cell Research* **1833**(7): 1573-1581.

Stewart EJ, Aslund F, Beckwith J. 1998. Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *The EMBO Journal* **17**(19): 5543-5550.

Stone SL, Anderson EM, Mullen RT, Goring DR. 2003. ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. *Plant Cell* **15**(4): 885-898.

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, Mori M, Hara-Nishimura I. 2010. Stomagen positively regulates stomatal density in *Arabidopsis*. *Nature* **463**(7278): 241-244.

Swanson WJ, Vacquier VD. 1997. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. *Proceedings of the National Academy of Sciences of the United States of America* **94**(13): 6724-6729.

Swanson WJ, Vacquier VD. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**(5377): 710-712.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**(2): 137-144.

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. 2000. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. *Nature* **403**(6772): 913-916.

Takayama S, Isogai A. 2005. Self-incompatibility in plants. *Annual Review of Plant Biology* **56**: 467-489.

Takayama S, Isogai A, Tsukamoto C, Ueda Y, Hinata K, Okazaki K, Suzuki A. 1987. Sequences of *S*-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. *Nature* **326**(6108): 102-105.

Takayama S, Shiba H, Iwano M, Asano K, Hara M, Che FS, Watanabe M, Hinata K, Isogai A. 2000. Isolation and characterization of pollen coat proteins of *Brassica campestris* that interact with *S* locus-related glycoprotein 1 involved in pollen-stigma adhesion. *Proceedings of the National Academy of Sciences of the United States of America* **97**(7): 3765-3770.

Takayama S, Shimosato H, Shiba H, Funato M, Che FS, Watanabe M, Iwano M, Isogai A. 2001. Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. *Nature* **413**(6855): 534-538.

Takebayashi N, Brewer PB, Newbigin E, Uyenoyama MK. 2003. Patterns of variation within self-incompatibility loci. *Molecular Biology and Evolution* **20**(11): 1778-1794.

Takeuchi H, Higashiyama T. 2012. A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *Plos Biology* **10**(12): e1001449.

Tang WH, Ezcurra I, Muschietti J, McCormick S. 2002. A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell* **14**(9): 2277-2287.

Tang WH, Kelley D, Ezcurra I, Cotter R, McCormick S. 2004. LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth *in vitro*. *Plant Journal* **39**(3): 343-353.

Terras FRG, Schoofs HME, Debolle MFC, Vanleuven F, Rees SB, Vanderleyden J, Cammue BPA, Broekaert WF. 1992. Analysis of 2 novel classes of plant antifungal proteins from radish (*Raphanus sativus* L) seeds. *Journal of Biological Chemistry* **267**(22): 15301-15309.

Tesfaye M, Silverstein KAT, Nallu S, Wang L, Botanga CJ, Gomez SK, Costa LM, Harrison MJ, Samac DA, Glazebrook J, *et al*. 2013. Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. *Plos One* **8**(3): e58992.

Teshima KM, Innan H. 2004. The effect of gene conversion on the divergence between duplicated genes. *Genetics* **166**(3): 1553-1560.

Thevissen K, Ghazi A, DeSamblanx GW, Brownlee C, Osborn RW, Broekaert WF. 1996. Fungal membrane responses induced by plant defensins and thionins. *Journal of Biological Chemistry* **271**(25): 15018-15025.

Thomas CL, Bayer EM, Ritzenthaler C, Fernandez-Calvino L, Maule AJ. 2008. Specific targeting of a plasmodesmal protein affecting cell-to-cell communication. *Plos Biology* **6**(1): 180-190.

Thomas SG, Franklin-Tong VE. 2004. Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature* **429**(6989): 305-309.

Tovar-Mendez A, Kumar A, Kondo K, Ashford A, Baek YS, Welch L, Bedinger PA, McClure BA. 2014. Restoring pistil-side self-incompatibility factors recapitulates an interspecific reproductive barrier between tomato species. *Plant Journal* **77**(5): 727-736.

Traverso JA, Pulido A, Rodriguez-Garcia MI, Alche JD. 2013. Thiol-based redox regulation in sexual plant reproduction: new insights and perspectives. *Frontiers in Plant Science* **4**:465.

Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert F-M, Vandermoere F, Morrice NA, Swift S, Rothbauer U, Leonhardt H, *et al*. 2008. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *Journal of Cell Biology* **183**(2): 223-239.

Trinquier G, Sanejouand YH. 1998. Which effective property of amino acids is best preserved by the genetic code? *Protein Engineering* **11**(3): 153-169.

Trotochaud AE, Jeong S, Clark SE. 2000. CLAVATA3, a multimeric ligand for the CLAVATA1 receptor-kinase. *Science* **289**(5479): 613-617.

Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R, Isokawa S, Pavlidis P, Stadler T, Suzuki G, Takayama S, Watanabe M, Shimizu KK. 2010. Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene. *Nature* **464**(7293): 1342-1346.

Twell D. 2006. A blossoming romance: gamete interactions in flowering plants. *Nature Cell Biology* **8**(1): 14-16.

Udagawa H, Ishimaru Y, Li F, Sato Y, Kitashiba H, Nishio T. 2010. Genetic analysis of interspecific incompatibility in *Brassica rap*a. *Theoretical and Applied Genetics* **121**(4): 689-696.

Updegraff EP, Zhao F, Preuss D. 2009. The extracellular lipase EXL4 is required for efficient hydration of Arabidopsis pollen. *Sexual Plant Reproduction* **22**(3): 197-204.

Vacquier VD. 1998. Evolution of gamete recognition proteins. *Science* **281**(5385): 1995-1998.

Vacquier VD, Swanson WJ, Lee YH. 1997. Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? *Journal of Molecular Evolution* **44**: S15-S22.

Vaddepalli P, Herrmann A, Fulton L, Oelschner M, Hillmer S, Stratil TF, Fastner A, Hammes UZ, Ott T, Robinson DG, *et al*. 2014. The C2-domain protein QUIRKY and the receptor-like kinase STRUBBELIG localize to plasmodesmata and mediate tissue morphogenesis in *Arabidopsis thaliana*. *Development* **141**(21): 4139-4148.

Valkonen JPT, Watanabe KN. 1999. Autonomous cell death, temperature sensitivity and the genetic control associated with resistance to cucumber mosaic virus (CMV) in diploid potatoes (*Solanum* spp.). *Theoretical and Applied Genetics* **99**(6): 996-1005.

Vallender EJ, Lahn BT. 2004. Positive selection on the human genome. *Human Molecular Genetics* **13**: R245-R254.

Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, *et al*. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**(5969): 1122-1126.

Vanoosthuyse V, Miege C, Dumas C, Cock JM. 2001. Two large *Arabidopsis thaliana* gene families are homologous to the *Brassica* gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. *Plant Molecular Biology* **46**(1): 17-34.

Veitia RA. 2005. Gene dosage balance: deletions, duplications and dominance. *Trends in Genetics* **21**(1): 33-35.

Vogler DW, Kalisz S. 2001. Sex among the flowers: the distribution of plant mating systems. *Evolution* **55**(1): 202-204.

von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**(6887): 399-403.

Wang L, Clarke LA, Eason RJ, Parker CC, Qi B, Scott RJ, Doughty J. 2016. PCP-B class pollen coat proteins are key regulators of the hydration checkpoint in *Arabidopsis thaliana* pollen–stigma interactions. *New Phytologist*. Published online September 6, 23016. http://dx.doi.org/10.1111/nph.14162.

Wang T, Liang L, Xue Y, Jia PF, Chen W, Zhang MX, Wang YC, Li HJ, Yang WC. 2016. A receptor heteromer mediates the male perception of female attractants in plants. *Nature* **531**(7593): 241-244.

Wendt T, Canela MBF, Klein DE, Rios RI. 2002. Selfing facilitates reproductive isolation among three sympatric species of *Pitcairnia* (Bromeliaceae). *Plant Systematics and Evolution* **232**(3-4): 201-212.

Weterings K, Russell SD. 2004. Experimental analysis of the fertilization process. *Plant Cell* **16**: S107-S118.

Wheeler MJ, de Graaf BHJ, Hadjiosif N, Perry RM, Poulter NS, Osman K, Vatovec S, Harper A, Franklin FCH, Franklin-Tong VE. 2009. Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature* **459**(7249): 992-995.

Wolters-Arts M, Lush WM, Mariani C. 1998. Lipids are required for directional pollen-tube growth. *Nature* **392**(6678): 818-821.

Wright S. 1939. The distribution of self-sterility alleles in populations. *Genetics* **24**(4): 538-552.

Wright SI, Barrett SCH. 2010. The long-term benefits of self-rejection. *Science* **330**(6003): 459-460.

Wright SI, Kalisz S, Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B-Biological Sciences* **280**: 20130133.

Wu JY, Wang S, Gu YC, Zhang SL, Publicover SJ, Franklin-Tong VE. 2011. Self-incompatibility in *Papaver rhoeas* activates nonspecific cation conductance permeable to Ca^{2+} and K^{+}. *Plant Physiology* **155**(2): 963-973.

Wu X, Cai G, Gong F, An S, Cresti M, Wang W. 2015. Proteome profiling of maize pollen coats reveals novel protein components. *Plant Molecular Biology Reporter* **33**(4): 975-986.

Wu Y, Li Q, Chen X-Z. 2007. Detecting protein-protein interactions by far western blotting. *Nature Protocols* **2**(12): 3278-3284.

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuehrer J, von Mering C, Grossniklaus U. 2010. *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current Biology* **20**(6): 506-512.

Yang H, Kaur N, Kiriakopolos S, McCormick S. 2006. EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L. *Planta* **224**(5): 1004-1014.

Yang ZH. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of Molecular Evolution* **51**(5): 423-432.

Yang ZH. 2007. PAML 4: PhylogeneticAnalysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**(8): 1586-1591.

Yang ZH, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* **15**(12): 496-503.

Yang ZH, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* **46**(4): 409-418.

Yang ZH, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**(1): 32-43.

Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**(4): 1107-1118.

Yasukawa T, Kaneiishii C, Maekawa T, Fujimoto J, Yamamoto T, Ishii S. 1995. Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. *Journal of Biological Chemistry* **270**(43): 25328-25331.

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, *et al*. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**(7378): 520-524.

Yount NY, Andres MT, Fierro JF, Yeaman MR. 2007. The gamma-core motif correlates with antimicrobial activity in cysteine-containing kallocin-1 originating from transferrins. *Biochimica Et Biophysica Acta-Biomembranes* **1768**(11): 2862-2872.

Yount NY, Yeaman MR. 2004. Multidimensional signatures in antimicrobial peptides. *Proceedings of the National Academy of Sciences of the United States of America* **101**(19): 7363-7368.

Zhang D, Liang W, Yin C, Zong J, Gu F, Zhang D. 2010. *OsC6*, Encoding a lipid transfer protein, is required for postmeiotic anther development in rice. *Plant Physiology* **154**(1): 149-162.

Zhang G, Sun YF, Li YM, Dong YL, Huang XL, Yu YT, Wang JM, Wang XM, Wang XJ, Kang ZS. 2013. Characterization of a wheat C2 domain protein encoding gene regulated by stripe rust and abiotic stresses. *Biologia Plantarum* **57**(4): 701-710.

Zhang JZ. 2003. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18**(6): 292-298.

Zhou P, Silverstein KAT, Gao L, Walton JD, Nallu S, Guhlin J, Young ND. 2013. Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinformatics* **14**: 335.

Appendix 1 Supplementary information for Chapter 3

New Phytologist Supporting Information Figs S1.1-S1.12, Tables S1.1-S1.4, and Methods S1.1

Article title: PCP-B class pollen coat proteins are key regulators of the hydration checkpoint in *Arabidopsis thaliana* pollen-stigma interactions

Authors: Ludi Wang, Lisa A Clarke, Russell J Eason, Christopher C Parker, Baoxiu Qi, Rod J Scott and James Doughty

Article acceptance date: 23 July 2016

The following Supporting Information is available for this article:

Fig. S1.1 Locations of T-DNA insertions

Fig. S1.2 RT-PCR analysis results of stage 12 anthers in *pcp-b* mutants

Fig. S1.3 N-terminal sequencing of two PCP-B proteins purified from *Brassica oleracea* pollen coat

Fig. S1.4 Phylogeny of 282 predicted PCP-B-like protein sequences

Fig. S1.5 RNA-RNA *in situ* hybridisation study of *AtPCP-Bγ* expression in *Arabidopsis thaliana* anthers.

Fig. S1.6 Histochemical staining for GUS activity driven by *AtPCP-Bα* and *AtPCP-Bδ* promoters in Arabidopsis tissues

Fig. S1.7 Pollen hydration profiles of wild-type and *pcp-b* triple mutant grains in a humid chamber

Fig. S1.8 Scanning Electron Microscopic (SEM) analysis of exine layer and pollen coat morphology

Fig. S1.9 Transmission Electron Microscopic (TEM) analysis of exine layer and pollen coat morphology

Fig. S1.10 Comparison of pollen tube growth for wild-type and *pcp-b* triple mutant plants

Fig. S1.11 Homologous alignments of ESF1.3 and AtPCP-Bs for protein structure predictions

Fig. S1.12 Predicted protein structure homology models of AtPCP-Bα, β and δ

Table S1.1 PCR primers used in this study

Table S1.2 Numbers and abbreviations of predicted PCP-B-like proteins in species and families

Table S1.3 Average seed count values of Arabidopsis wild-type and *pcp-b* mutants

Table S1.4 Statistics for AtPCP-B protein structural predictions

Method S1.1 Histochemical staining for β-glucuronidase (GUS) activity

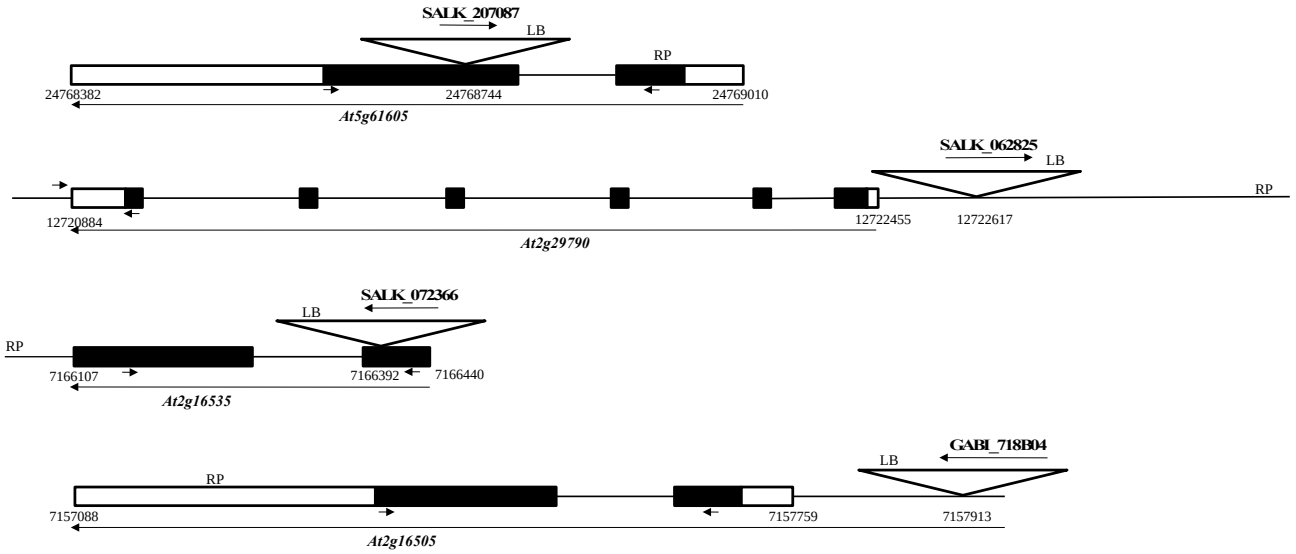


Fig. S1.1 T-DNA insertion locations of the *pcp-b* mutants. White boxes, UTR untranscribed regions; black boxes, exons; horizontal lines, introns; triangles, T-DNA insertion sites. LB and RP indicate the location of primers used in the confirmation of T-DNA insertions. Short arrows indicate the locations of primers used in RT-PCR for the conformation of gene knockouts. *PCP-B* mutant lines are named as follows: SALK_207087, *pcp-bα-1*; SALK_062825, *pcp-bβ-1; SALK_072366, pcp-bγ-1; GABI_718B04, pcp-bδ-1*.

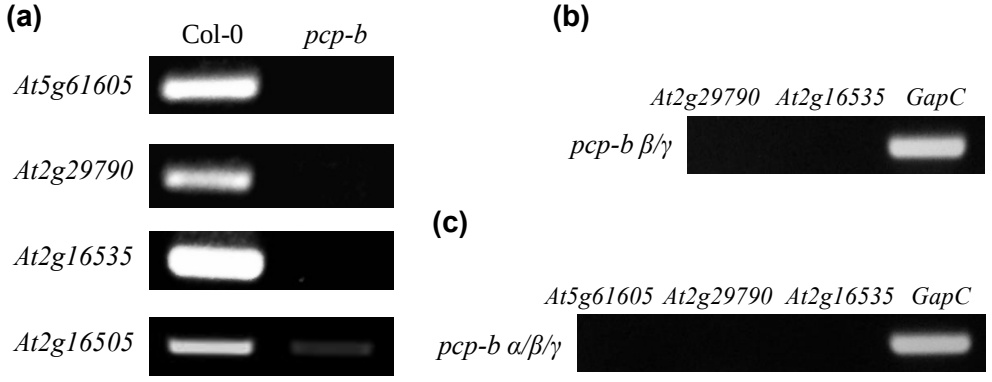


Fig. S1.2 RT-PCR expression analysis of *PCP-B*s in stage 12 anthers of T-DNA single gene mutant lines (a), the *pcp-b β/γ* double gene mutant line (b) and *pcp-b α/β/γ* triple gene mutant line (c).

PCP-B1: AGNAAKPT**C**KQTP**C**HELKPNHT**C**S**C**

PCP-B2: AGNAAKQT**C**KQMN**C**DTGDKN

Fig. S1.3 N-terminal sequencing of two PCP-B proteins purified from *Brassica oleracea* pollen coat. The conserved cysteine residues are shown as bold letters. The shared six amino acid N-terminal region is underlined.

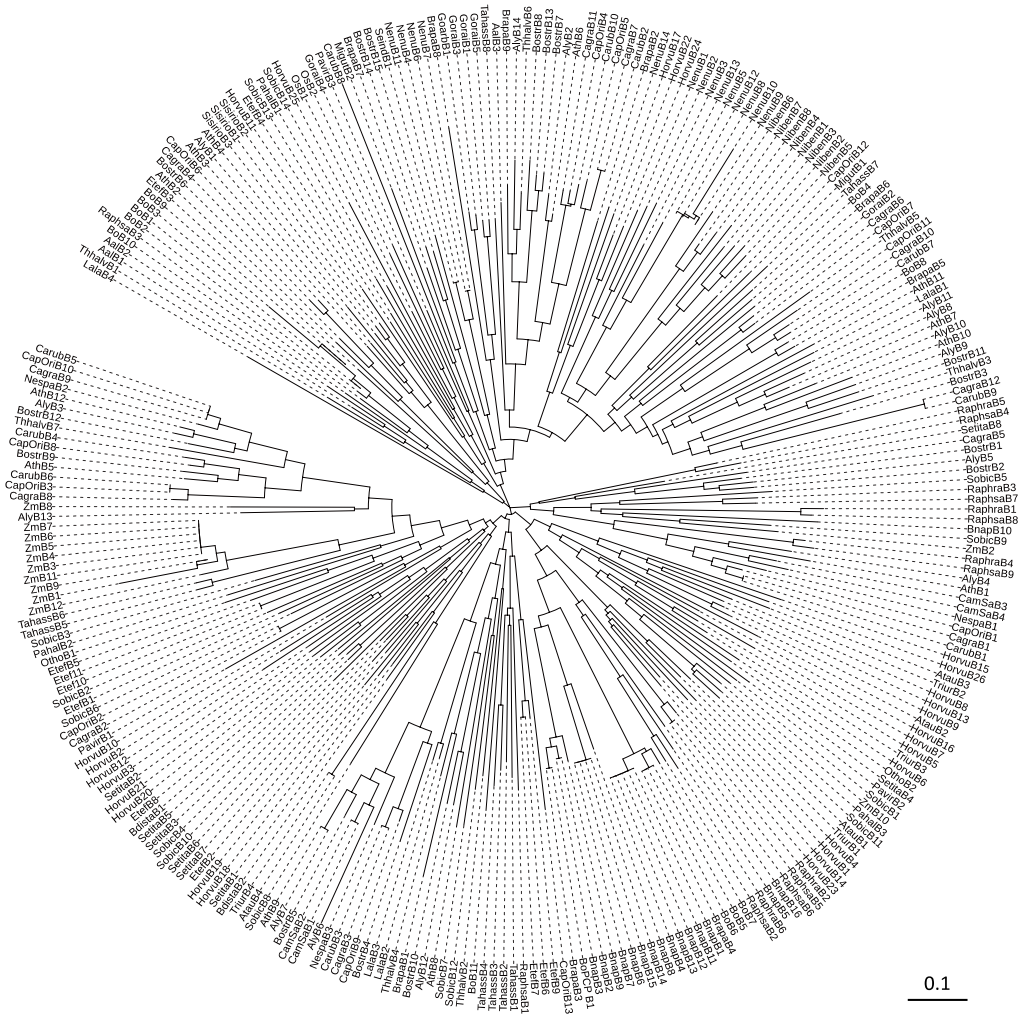


Fig. S1.4 Phylogeny of PCP-Bs and PCP-B-like proteins in angiosperms. The Neighbour-Joining Tree was constructed using amino acid sequences of the predicted mature proteins. Branch length is scaled to the scale bar defined as 0.1 substitutions per site. Abbreviated gene names can be found in Table S1.2.

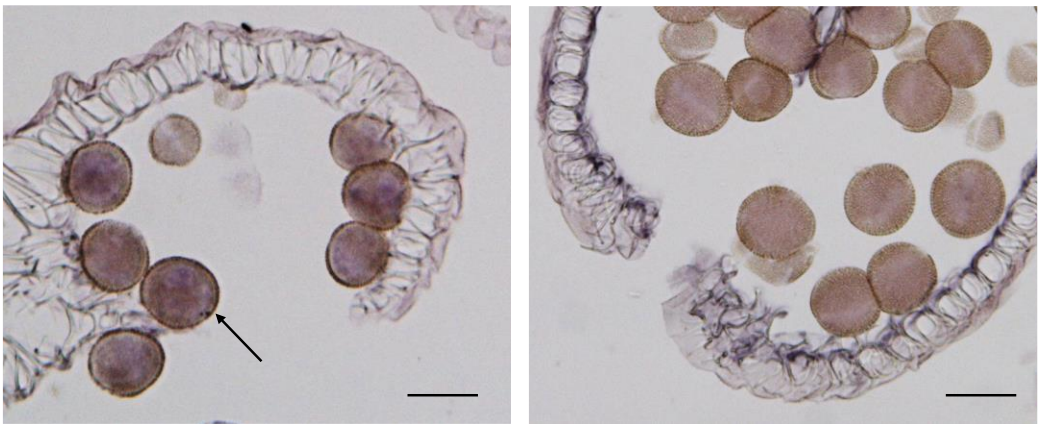


Fig. S1.5 RNA-RNA *in situ* hybridisation study of *AtPCP-Bγ* expression in *Arabidopsis thaliana* anthers. Image at left - transverse anther section treated with an antisense (+ve) *AtPCP-Bγ* DIG-labelled riboprobe, a clear signal (arrow) is observed within the majority of pollen grains. Image at right – transverse anther section treated with a control 'sense' (-ve) riboprobe with no signal being detectable in pollen grains. Scale bar: 10 μm.

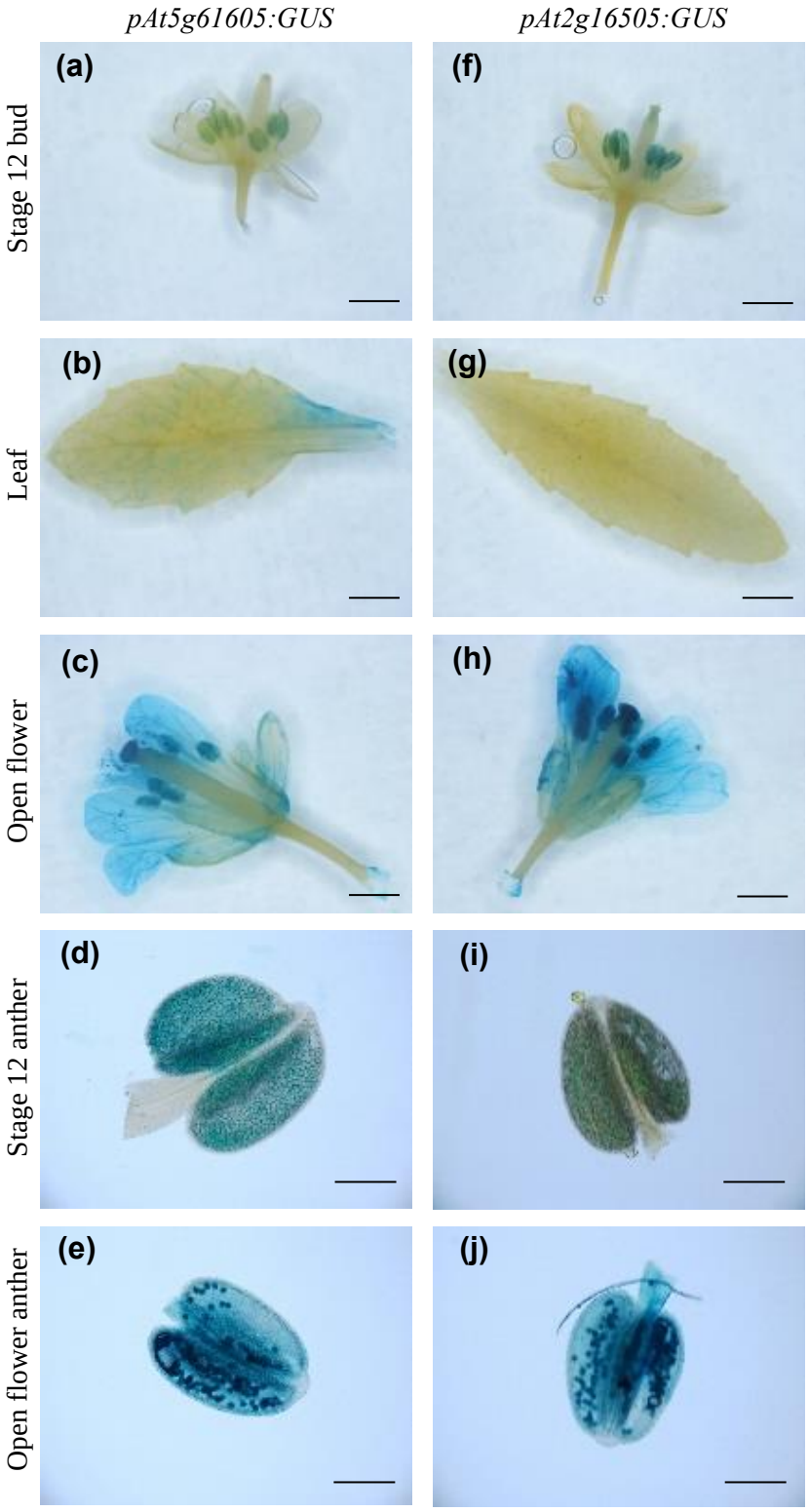


Fig. S1.6 GUS expression in *Arabidopsis thaliana* tissues driven by *AtPCP-Bα* promoter *pAt5g61605* (a-e) and *AtPCP-Bδ* promoter *pAt2g16505* (f-j). Scale bars: 1 mm (a-c and f-h); 0.2 mm (d, e, i and j). Signal in stigmatic tissues is the result of pollen deposition.

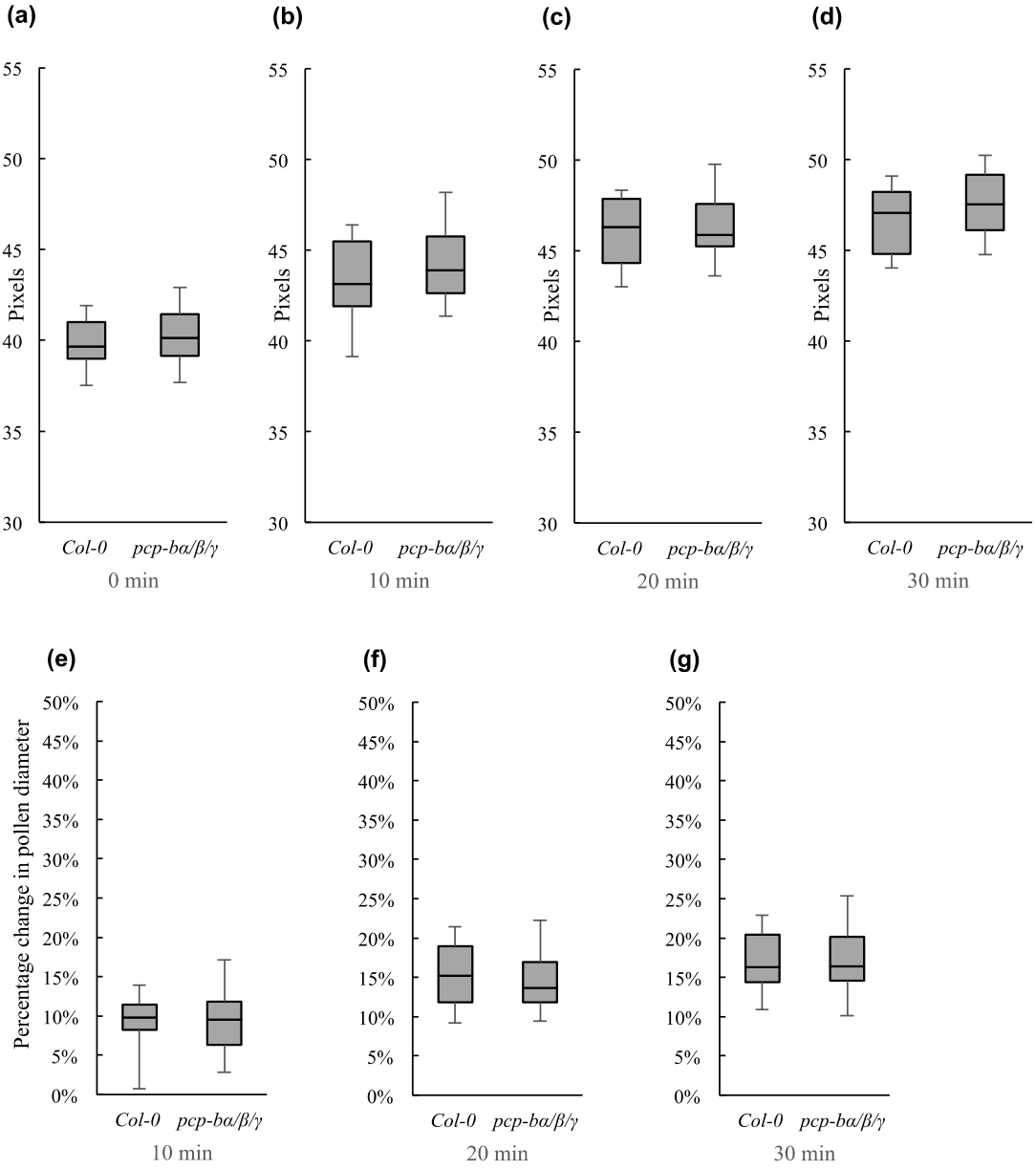


Fig. S1.7 Pollen hydration profiles of wild-type and *pcp-b* triple mutant pollen in a humid chamber. **a-d**, Pollen diameter distributions at 0min, 10min, 20min, 30min of hydration in a humid chamber. Box plots depict the 25% quartile, median, 75% quartile and full range of values. Sample sizes: 20. **e-g**, distribution of percentage change in pollen diameter at 10min, 20min, 30min of hydration in the humid chamber. Box plots depict the 25% quartile, median, 75% quartile and full range of values. Sample sizes: 20. *P*-values: a, 0.32; b, 0.30; c, 0.79; d, 0.13; e, 0.71 f, 0.46; g, 0.52 (Welsh's *t*-test). 1.8 pixels = 1 μm.

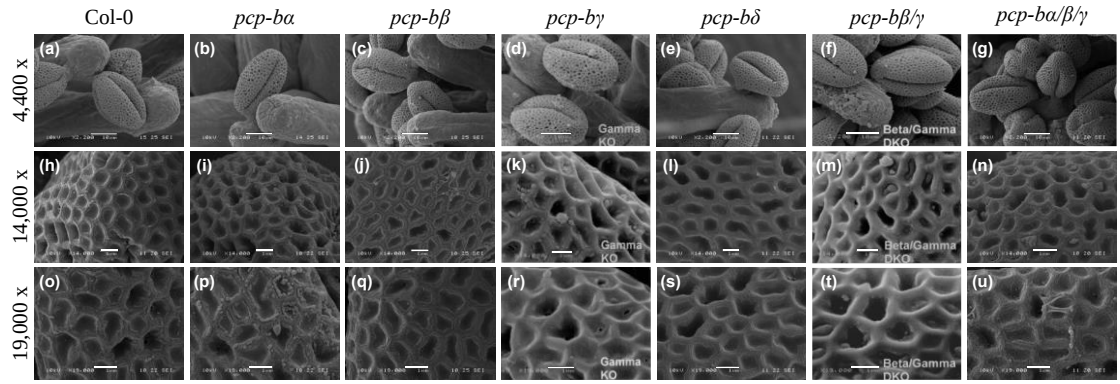


Fig. S1.8 Scanning Electron Microscopic (SEM) analysis of exine layer and pollen coat morphology. Scale bar: a-g, 10µm. h-u, 1µm. a, h and o, wild-type Col-0. b, i and p, *pcp-bα*. c, j and q, *pcp-bβ*. d, k and r, *pcp-bγ*. e, l and s, *pcp-bδ*. f, m and t, *pcp-bβ/γ*. g, n and u, *pcp-bα/β/γ*.

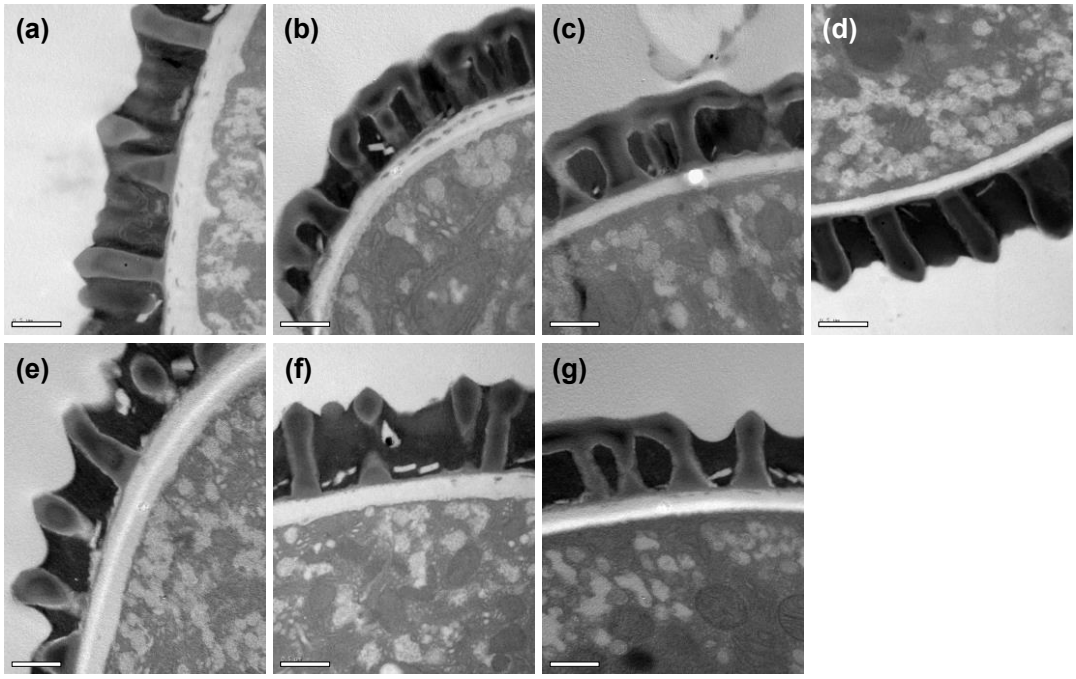


Fig. S1.9 Transmission Electron Microscopic (TEM) analysis of exine layer and pollen coat morphology. Scale bar: 0.5 µm. a, wild-type Col-0. b, *pcp-bα*. c, *pcp-bβ*. d, *pcp-bγ*. e, *pcp-bδ*. f, *pcp-bβ/γ*. g, *pcp-bα/β/γ*.

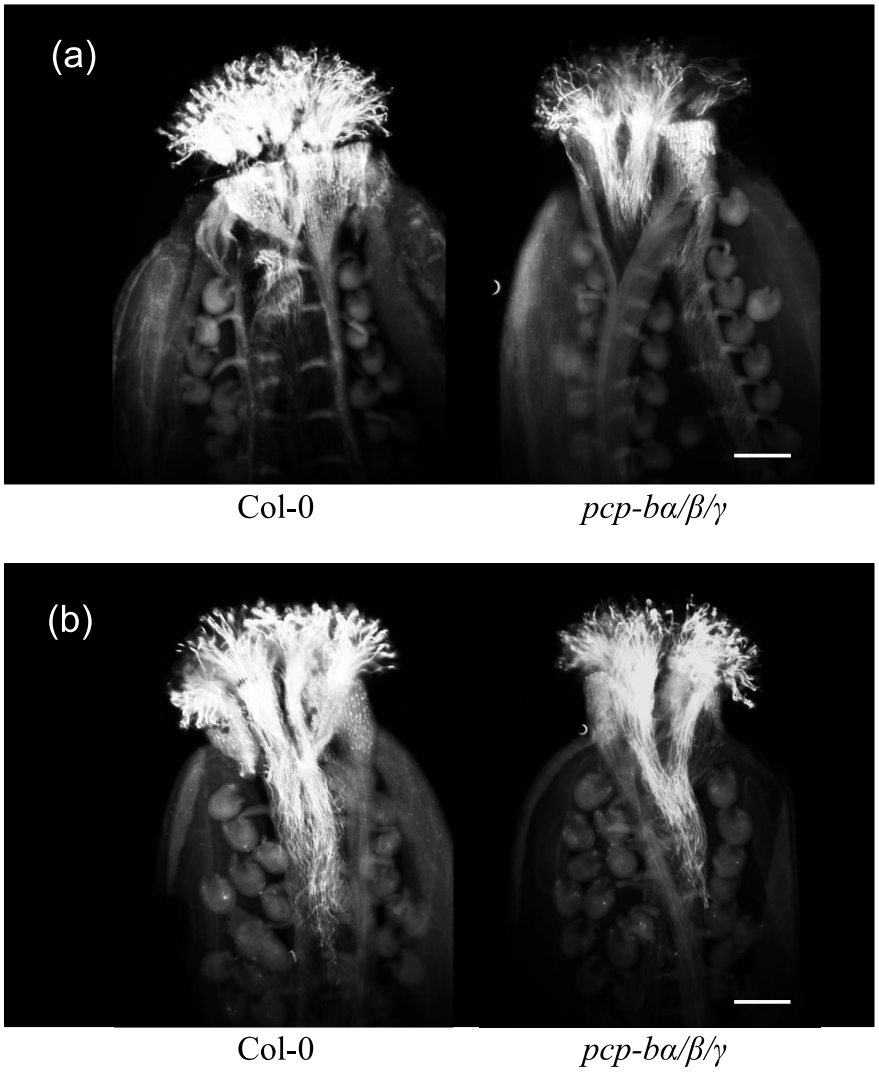


Fig. S1.10 Pollen tube growth comparisons of wild-type and *pcp-b* triple mutant after two hours (a) and four hours (b) of pollination. Scale bar = 0.2 mm.

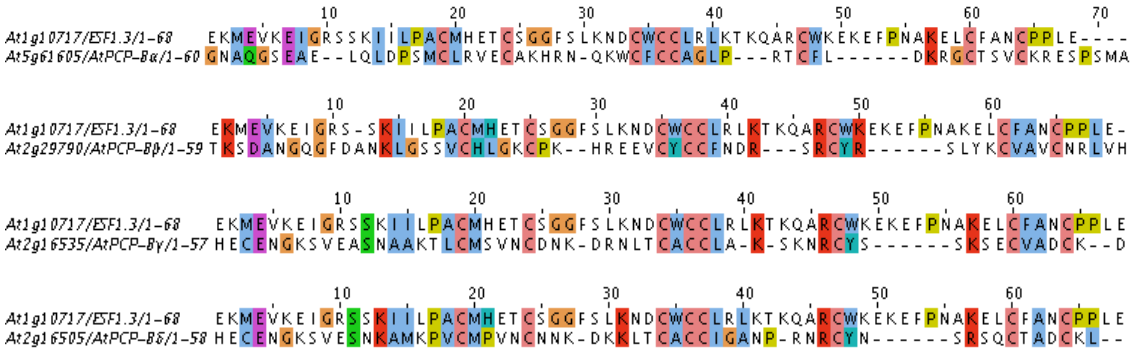


Fig. S1.11 Homologous alignments of ESF1.3 and AtPCP-Bs for protein structural predictions. Colour coding follows the default output for Clustal X (http://www.jalview.org/help/html/colourSchemes/clustal.html).

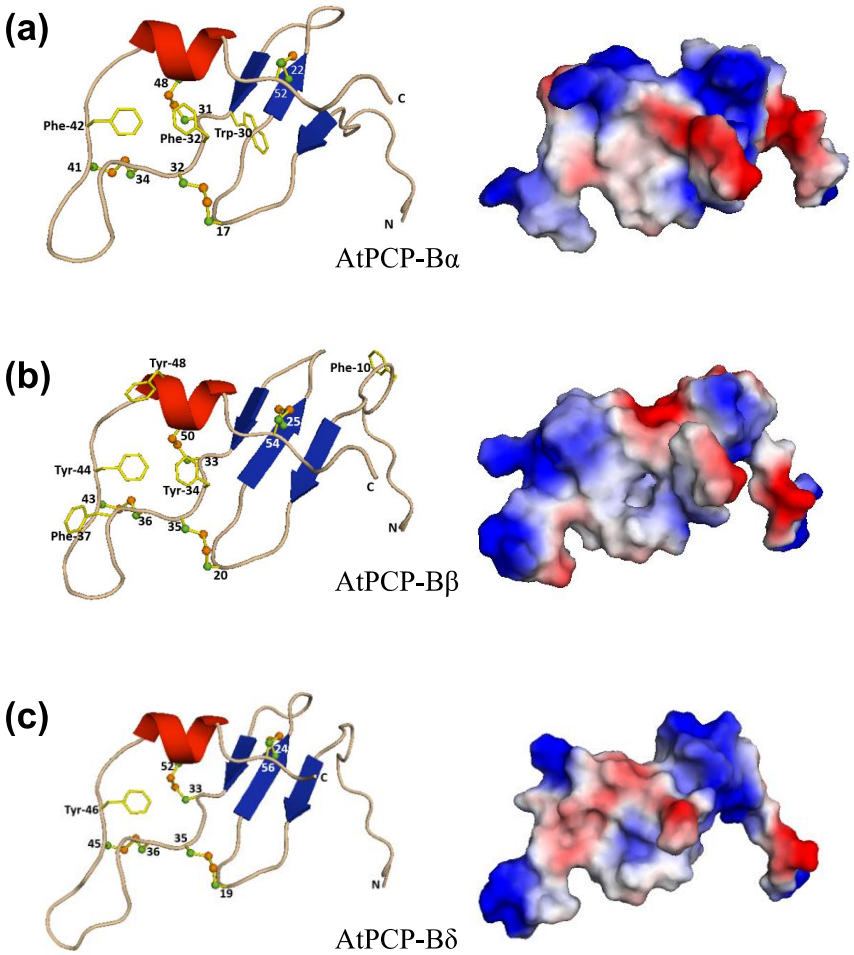


Fig. S1.12 Structural prediction models and surface distributions of electrostatic potential for AtPCP-Bs. AtPCP-Bα (a), AtPCP-Bβ (b) and AtPCP-Bδ (c) predicted structures are displayed with both disulphide bonds and aromatic residues being indicated.

Table S1.1 PCR primers used for *PCP-B* homologous gene expression and T-DNA line genotyping (a), *AtPCP-B* temporal gene expression pattern by RT-PCR (b) and genotyping of T-DNA mutant lines (c).

(a)

Amplicon	Forward primer (5'-3')	Reverse primer (5'-3')
GapC (At3g04120)	TCAGACTCGAGAAAGCTGCTAC	GATCAAGTCGACCACACGG
At5g61605	TGCTATCCTTTGCCTCTTCATG	AGCCATTGAAGGAGATTCGCG
At2g29790	CAAGGTTTTGACGCGAACAAG	ATGAACAAGCCGGTTGCAG
At2g16535	CGCATCATCCTCATTTCTTC	ACACCTATTCTTGGATTTGGCC
At2g16505	GGGCCTTTGTATCATCCTG	TTGCAGTCAGCCGTACATTG
At2g41415	CCCATGCATATCTCGTCTGC	CAGCAATAACAGCCTCCGTC
At2g16225	TGTTCTCTCTCTTCGCTCTACA	CCGAAACAACACCAACAATG
At1g27135	TCCCTCTTCGCTATGCATGA	TCGAAGCAGCACCAACAATC
At5g50345	GCATCATCCTCGTTTCATTG	GGGTTAAGTCTAAGACACTC
At4g15953	TTCGTTTCCCTCTTTGGTGTGC	GGCCCTAATCCTTTGCTTACTC
At1g10747	ACTCATGCTCTCTCTCGTCG	CTTCCCAGCAAAGATCTGGC
At1g10745	CACAAACAGTTCTCATCTCC	AAACAACACCAACAATCCCG
At1g10717	ATCATGCTCTCCCTCTTTGCTC	GGCGAAACAAAGCTCCTTAG

(b)

Amplicon	Forward primer (5'-3')	Reverse primer (5'-3')
AtPCP-Bα (At5g61605)	TGCCTCTTCATGATTTTCCTCGTTCC	TTTATCAAGAAAACAGGTCCTCGG
AtPCP-Bβ (At2g29790)	GTAGTTTCTCTCGTTCCTCATGG	CGGTTGCAGACAGCCACACAC
AtPCP-Bγ (At2g16535)	CATCATCCTCATTTCTTCATTTCC	TTGCAGTCAGCAACACATTCTG
AtPCP-Bδ (At2g16505)	CATCCTGATTTCTTTCTTTCCTCTTC	GCCGTACATTGTGATCTGCTATTG
GapC (At3g04120)	CACTTGAAGGGTGGTGCCAAG	CCTGTTGTCGCCAACGAAGTC

(c)

Amplicon	Forward primer (5'-3')	Reverse primer (5'-3')
SALK_207087	GTTATGCCAATTCCAAAAGGC	TGCCTCTTCATGATTTTCCTC
SALK_062825	TTGAAATCCGAACCTGATTTG	TCTTATGGGGTTTTTGTGCAG
SALK_072366	TCCGTGGACTTGTGGTATACC	TTTCTTAATTCCTTAGTGGAGCTTG
GABI_718B04	TGGGACAGATTAAGAAGTTACGG	TGAAAACTCGTAGACCGCAAC
SALK line insertion left border	ATTTTGCCGATTTCGGAAC	-
GABI line insertion left border	ATAATAACGCTGCGGACATCTACATTTT	-

Table S1.2 Numbers and abbreviations of predicted PCP-B-like proteins in species and families.

Family	Species	Protein names
Brassicaceae 160	*Arabidopsis thaliana* 12	AthB1-12
	Arabidopsis lyrata 14	AlyB1-14
	Capsella rubella 10	CarubB1-10
	Capsella grandiflora 12	CagraB1-12
	Capsella orientalis 13	CapOriB1-13
	Neslia paniculata 3	NespaB1-3
	Camelina sativa 4	CamSaB1-4
	Leavenworthia alabamica 4	LalaB1-4
	Boechera stricta 15	BostrB1-15
	Arabis alpina 3	AalB1-3
	Brassica oleracea 12	BoB1-11, BoPCP-B1
	Brassica rapa 9	BrapaB1-9
	Brassica napus 16	BnapB1-16
	Raphanus raphanistrum 6	RaphraB1-6
	Raphanus sativus 9	RaphsaB1-9
	Sisymbrium irio 3	SisirioB1-3
	Eutrema salsugineum 7	ThhalvB1-7
	Tarenaya hassleriana 8	TahassB1-8
Malvaceae 6	*Gossypium raimondii* 5	GoraiB1-5
	Gossypium arboreum 1	GoarbB1
Poaceae 91	*Oryza sativa* 2	OsB1-2
	Hordeum vulgare 26	HorvuB1-26
	Triticum urartu 4	TriurB1-4
	Aegilops tauschii 4	AtauB1-4
	Brachypodium distachyon 2	BdistaB1-2
	Zea mays 12	ZmB1-12
	Sorghum bicolor 14	SobicB1-14
	Panicum virgatum 3	PavirB1-3
	Panicum hallii 3	PahalB1-3
	Setaria italica 8	SetitaB1-8
	Oropetium thomaeum 2	OthoB1-2
	Eragrostis tef 11	EtefB1-11
Nelumbonaceae 14	*Nelumbo nucifera* 14	NenuB1-14
Pedaliaceae 1	*Sesamum indicum* 1	SeindB1
Solanaceae 8	*Nicotiana benthamiana* 8	NibenB1-8
Phrymaceae 2	*Mimulus guttatus* 2	MigutB1-2

Table S1.3 Mean seed count values for Arabidopsis wild-type and *pcp-b* mutants.

Genotype	Mean number of seeds per silique±standard deviation (n=5)	*p*-value (Welsh's *t*-test)
Wile-type	60±3	-
pcp-bα	62±4	0.35
pcp-bβ	64±5	0.92
pcp-bγ	55±4	0.88
pcp-bδ	56±4	0.92
pcp-bβ/γ	61±5	0.98
pcp-bα/β/γ	62±3	0.96

Table S1.4 The Global Model Quality Estimation (GMQE) and Qualitative Model Energy Analysis (QMEAN4) scores of the predicted AtPCP-B protein models.

Protein model	Sequence similarity	GMQE	QMEAN4
AtPCP-Bα	0.32	0.59	-5.52
AtPCP-Bβ	0.34	0.65	-4.74
AtPCP-Bγ	0.36	0.69	-6.14
AtPCP-Bδ	0.36	0.69	-5.37

Method S1.1 Histochemical staining for β-glucuronidase (GUS) activity

Leaves, open flowers and stage 12 buds of *pAt5g61605: GUS* and *pAt2g16505: GUS* lines were transferred into GUS substrate solution (100mM sodium phosphate buffer pH7.0, 10mM EDTA, 0.1% v/v Triton X-100, 1mM potassium ferricyanide, 2mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronide) and vacuum-infiltrated for 5 minutes before overnight incubation at 37°C. Samples were then transferred into 50% (v/v) ethanol to remove chlorophyll and mounted in 50% (v/v) glycerol for imaging.

Appendix 2 Supplementary information for Chapter 4

2.1 Supporting figures and tables

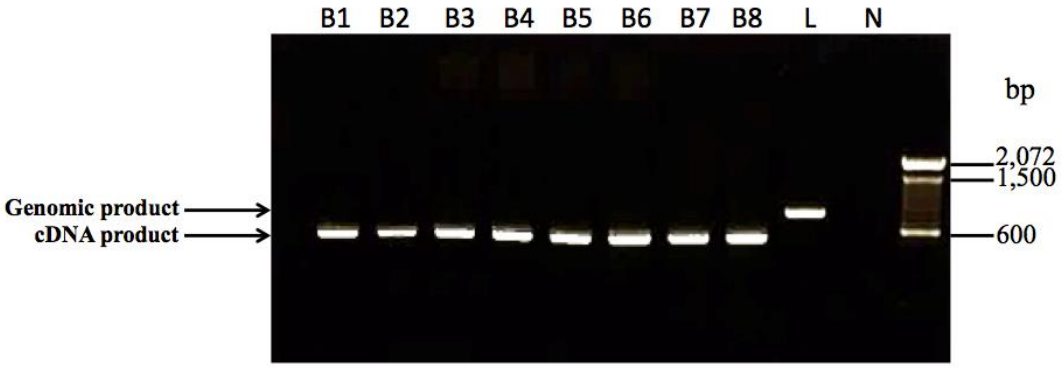


Figure S4.1 | PCR products amplified from anther cDNA and genomic DNA from leaves by using GapC gene specific primers. B1 to B8, the PCR products from different samples with stage 12 anther cDNA synthesis products with *GapC* specific primers; L, the PCR product with leaf genomic DNA and *GapC* specific primers; N, negative control without DNA template.

Table S4.1 | The primers used for AtPCP-B gene amplification

Gene loci	Primers	Primer sequences
At5g61605	AtPCP-Bα F	5'-ttatatccatgggctctgaggcggagctgcaattag
At5g61605	AtPCP-Bα R	5'-tataactcgagagccattgaaggagattcgcgc
At2g29790	AtPCP-Bβ F	5'-ttatatccatgggccaaggttttgacgcgaacaag
At2g29790	AtPCP-Bβ R	5'-tataactcgagatgaacaagccggttgcagac
At2g16535	AtPCP-Bγ F	5'-ttatatccatgggcgaaaatggaaaaagtgttgaagcg
At2g16535	AtPCP-Bγ R	5'-tataactcgagatctttgcagtcagcaacacattc
At2g16505	AtPCP-Bδ F	5'-ttatatccatgggcgaaaatggaaaaagtgttgaatcg
At2g16505	AtPCP-Bδ R	5'-tataactcgagaagtttgcagtcagccgtacattg

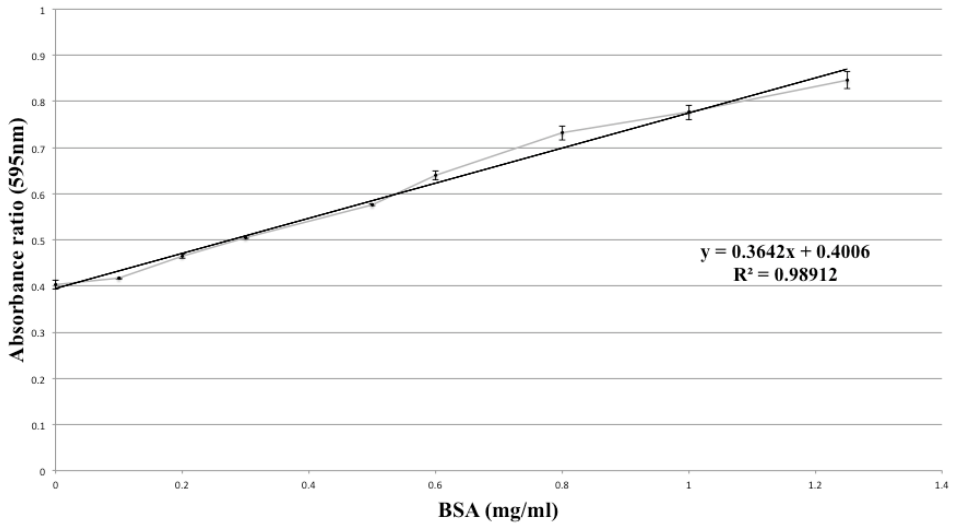


Figure S4.2 | The standard curve of Bradford Assay carried out with BSA.

Table S4.2 | Evaluation of protein concentration by Bradford assay

Samples	A_{595}	Concentration (mg/ml)
AtPCP-Bα Elution 1	0.492	0.251
AtPCP-Bα Elution 2	0.456	0.152
AtPCP-Bα Elution 3	0.408	0.020
AtPCP-Bβ Elution 1	0.442	0.114
AtPCP-Bβ Elution 2	0.438	0.103
AtPCP-Bβ Elution 3	0.409	0.023
AtPCP-Bγ Elution 1	0.459	0.160
AtPCP-Bγ Elution 2	0.438	0.103
AtPCP-Bγ Elution 3	0.403	0.007
AtPCP-Bδ Elution 1	0.439	0.105
AtPCP-Bδ Elution 2	0.440	0.108
AtPCP-Bδ Elution 3	0.405	0.012

2.2 Composition of solutions, buffers and media

2.2.1 Bacteria growth medium

LB medium (per liter)

Tryptone 10g

Yeast extract 5g

NaCl 10g

LB plates with carbenicillin

Agar 15g

LB medium 1 liter

Autoclave and cool to 50°C

Add 50mg/ml carbenicillin stock to a final concentration of 50µg/ml

Pour 20-30ml of medium into 85mm petri dishes and wait for harden

Store at 4°C for up to 1 month

2.2.2 TAE buffer (10 x, 1 liter)

48.4g Tris
20ml of 0.5M EDTA (pH 8.0)
11.44ml glacial acetic acid

2.2.3 SDS-PAGE buffer system

Gel preparation solutions

Tris-glycine gel

Solutions	Stacking gel 4%	Resolving gel 7.5%	12%	X%
40% acrylamide/bis 37.5:1	0.5 ml	1.88 ml	3 ml	2.5(X%) =A ml
0.5 M Tris-HCl pH 6.8	1.26 ml	-	-	-
1.5 M Tris-HCl pH 8.8	-	2.5 ml	2.5 ml	2.5 ml
10% SDS	50 µl	100 µl	100 µl	100 µl
ddH_20	3.18 ml	5.47 ml	4.35 ml	7.35-A ml
TEMED	5 µl	5 µl	5 µl	5 µl
10% APS	25 µl	50 µl	50 µl	50 µl
Total volume	5.02 ml	10.005 ml	10.005 ml	10.005 ml

2x sample buffer

125 mM Tris pH6.8
4% SDS
20% glycerol
0.01% Bromophenol blue
5% beta-mercaptoethanol (add fresh before use)

10x Electrode (Running) Buffer, pH 8.3

30.3 g Tris base (15 g/l)
144.0 g Glycine (72 g/l)

Tris-tricine gel

Solutions	Stacking gel 4%	Resolving gel 16.5%
40% acrylamide/bis 19:1	0.4 ml	4.125 ml
3 x gel buffer	1.33 ml	3.3 ml
50% glycerol	-	1 ml
10% APS	50 µl	50 µl
ddH_20	2.21 ml	1.515
TEMED	10 µl	10 µl
Total volume	4 ml	10 ml

3x gel buffer (Tris Cl/SDS, pH=8.45)

Dissolve 182 g Tris base (3.0 M) in 300 ml H2O. Adjust pH to 8.45 with 1 M HCl. Add H_2O to 500 ml. Filter solution through 0.45 µm filter. Dissolve 1.5 g SDS (0.3% W/V).

Store at 4° C.

1.5x Tricine Sample Buffer

2 ml 4x Tris Cl/SDS, pH=6.8 (0.08M)

4.0 ml (5.0 g) glycerol (40% V/V)

2.0 ml 10% SDS (2% V/V)

0.2 ml β-mercaptoethanol

0.8 ml 0.5% Coomassie blue R-250

Add water to 10 ml and mix

10x Tris/Tricine/SDS Reservoir/Running Buffer

Dissolve 60.55 g Tris base (1.0 M) and 89.6 g tricine (1.0M) in 400 ml H_2O. Add 5.0 g SDS (1.0%) Do not adjust pH. Add H_2O to 500 ml.

Running Conditions:

Run gel at 100 V (constant) for 100 minutes. The current for two minigels should be about 65 mA. Use 1x Running Buffer.

Coomassie staining buffer (1liter)

Coomassie R250 1g
Glacial acetic acid 100ml
Methanol 250ml
ddH_2O 650ml

Protein gels were stained in this buffer for at least 30 minutes

Coomassie distaining buffer (1liter)

Glacial acetic acid 100ml
Methanol 250ml
ddH_2O 650ml

Protein gels were distained in fresh buffer for 3x 60 minutes

2.2.4 Western blot buffer system

Transfer buffer (1liter)
Tris 3g
Glycine 14.4g
Methanol 150ml
Add ddH_2O to 1 liter

Blocking buffer pH7.4 (1liter)
Casein 3g
NaCl 29.2g
Tris 2.4g

Adjust pH to 7.4 with HCl after dissolve. Add 4ml 25 Tween 20 (to 0.1)

2.2.5 Buffer system for protein purification with His-trap FF column

Binding/washing buffer:

20mM sodium phosphate (pH 7.4)

0.5M NaCl

10mM imidazole (the previous concentration was 20mM)

Elution buffer:

20mM sodium phosphate (pH 7.4)

0.5M NaCl

300mM imidazole

2.2.6 Buffer system for protein purification with His-Select® HF Nickel Affinity Gel

Binding/washing buffer:

50mM sodium phosphate (pH 8.0)

0.3M NaCl

10mM imidazole (the previous concentration was 20mM)

Elution buffer:

50mM sodium phosphate (pH 8.0)

0.3M NaCl

250mM imidazole

2.2.7 Protein cleavage

Purified protein dialysis and verification

- The purified protein solutions were dialysed in buffer for enterokinase cleavage (EK buffer):

  20mM Tris-HCl pH 8.0

  200mM NaCl (The concentration was optimised for maintaining the protein solubility)

  2mM $CaCl_2$
- The dialysed protein samples were verified on SDS-PAGE

The purified protein solutions were dialysed in buffer for enterokinase cleavage (EK buffer):

20mM Tris-HCl pH 8.0

200mM NaCl (The concentration was optimised for maintaining the protein solubility)

2mM $CaCl_2$

- The enterokinase solution was added to the fusion protein solution:

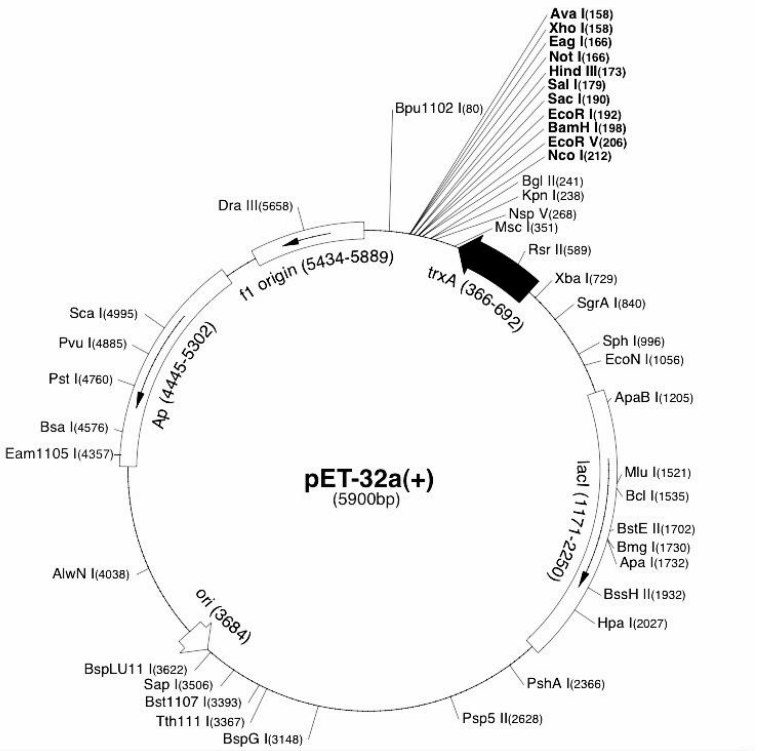
  0.5µl enterokinase (2µg/ml) was added to 4ml AtPCP-Bγ solution (64.25µg/ml)

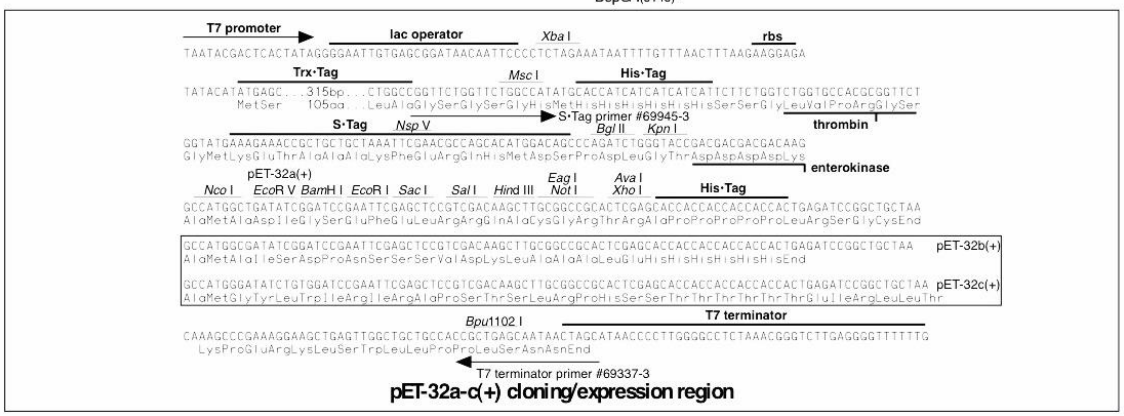
  0.3µl enterokinase (2µg/ml) was added to 4ml AtPCP-Bδ solution (28.56µg/ml)
- The reactions were incubated at room temperature for 16 hours. The cleavage products were verified on SDS-PAGE gel (Tris/glycine).

2.3 pET-32a vector map

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam* H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam* H I at 198 except for *Eco*R V, which cuts at 209.





2.4 Raw data of sequencings and alignments

by BLAST® (http://blast.ncbi.nlm.nih.gov)

AtPCP-Bα in pET-32a

NNNNNNNNGNNNNNTNCNNCTNNANNNNTTTTGTTTAACTTTAAGAAGGAGATATACATATGAGCGATAA
AATTATTCACCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTCGATT
TCTGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAG
GGCAAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGCATCCGTG
GTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGGT
CAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTCTGGTTCTGGCCATATGCACCATCATCATCATCAT
TCTTCTGGTCTGGTGCCACGCGGTTCTGGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACAT
GGACAGCCCAGATCTGGGTACCGACGACGACGACAAGGCCATGGGCTCTGAGGCGGAGCTGCAATTAGAT
CCATCAATGTGTCTGCGTGTCGAATGTGCAAAACACAGAAATCAAAAATGGTGTTTTTGCTGTGCCGGACT
ACCGAGGACCTGTTTTCTTGATAAACGAGGCTGTACGTCTGTCTGCAAGCGCGAATCTCCTTCAATGGCTCT
CGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCT
GCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAA
AGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGC
TTTACGGCNCCTCGACCCCAAAAAACTTGATTAGGGNGATGGTTCACGTAGTGNCCATCGCCCTGATAGAN
NNTTTTCGCCCTTTGACGTTGGAGTCCACGTNCTTTANANNGGANNNCTNNNTCCAAANNGNNANNNTNNA
CCCTATCTCGGNCNNTNNTTTGATTNNANGGATTTGCNATTCGNCTANNGNNAAAANGNNNTGANTTACAA
NNTNACGCNANTNANAANNTANNNTTACNNNNGNNNNNTNNNGGAAANNNCCGANCCNNTNNNNNTTNN
NNNNNNANTNNNNNNNNNNNNACNNNNNNNNNNNNNN

AtPCP-Bα sequencing data alignment

Arabidopsis thaliana chromosome 5, complete sequence

Sequence ID: ref|NC_003076.8| **Length:** 26975502 **Number of Matches:** 1

Range 1: 24768619 to 24768783 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
305 bits(165)	1e-81	165/165(100%)	0/165(0%)	Plus/Minus

Features: hypothetical protein

```
Query  539       TCTGAGGCGGAGCTGCAATTAGATCCATCAATGTGTCTGCGTGTCGAATGTGCAAAACAC  598
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  24768783  TCTGAGGCGGAGCTGCAATTAGATCCATCAATGTGTCTGCGTGTCGAATGTGCAAAACAC  24768724

Query  599       AGAAATCAAAAATGGTGTTTTTGCTGTGCCGGACTACCGAGGACCTGTTTTCTTGATAAA  658
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  24768723  AGAAATCAAAAATGGTGTTTTTGCTGTGCCGGACTACCGAGGACCTGTTTTCTTGATAAA  24768664

Query  659       CGAGGCTGTACGTCTGTCTGCAAGCGCGAATCTCCTTCAATGGCT  703
                 |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  24768663  CGAGGCTGTACGTCTGTCTGCAAGCGCGAATCTCCTTCAATGGCT  24768619
```

AtPCP-Bβ in pET-32a

NNNNNNNNNNNNNNNNNTCTANNNNNNTTTGTTNNNCTTTAAGAAGGAGATATACATATGAGCGATAAAA
TTATTNNNCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTCGATTTC
TGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGG
CAAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGCATCCGTGGTA
TCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGGTCAG
TTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTCTGGTTCTGGCCATATGCACCATCATCATCATCATTCT
TCTGGTCTGGTGCCACGCGGTTCTGGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGA
CAGCCCAGATCTGGGTACCGACGACGACGACAAGGCCATGGGCCAAGGTTTTGACGCGAACAAGTTAGGT
TCATCAGTGTGCCATCTTGGCAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCG
TAGCAGGTGTTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATCTCGAGCACCACCA
CCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGANNTGGCTGCTGCCACCGCTGAG
CAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCNTGAGGGGTTTTTTTGCTGAAAGGAGGAACTA
TATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGNGGTTACGCGC
AGCGNGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCNTCCCTTCCTTTCTCGCCACG
TTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCTTTAGGGTCCCGANTTAGTGNNTTACGGCACC
NCNACCCCAAAAACTTGATTANGGTGANGGTTCACGTANNGGGCCATCNCCCTGATAGACGGTTNTNNGC
NNTTGACGTTGGAGNCCACNTNNNTTAATANNNNNNTCTNNTTCAAANNNNNNNCANNCTCANCNTTANN
NNGNNCNNTTNNTTTTGATNTNNNNNGATTTNNCNANTNNNNNNNTNNTNNNNNAAAAAATNNNNCTGAN
TTAANNAAANNNNNNNCNNTTNNNNNAANTNTTTNNNNNNTTANNTTNGNGGNNCTTTTNNNNNNNNNGN
GNNNNNANNNNNNNNNNNNNTTNNNNNANNNNCNTNNANNTNNNNNNCNNCNNNNNNNANNNNNANAC
NCNNNNNNNANNNNNNGCNN

***AtPCP-Bβ* sequencing data alignment**

Arabidopsis thaliana chromosome 2, complete sequence

Sequence ID: ref|NC_003071.7| **Length:** 19698289 **Number of Matches:** 1

Range 1: 12720854 to 12721012 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
289 bits(156)	1e-76	159/160(99%)	1/160(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  537       GGGCCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGGCAAATGTCC  596
                  ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721012  GGG-CAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGGCAAATGTCC  12720954

Query  597       AAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTGTTATCGTAG  656
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720953  AAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTGTTATCGTAG  12720894

Query  657       TTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCAT  696
                  ||||||||||||||||||||||||||||||||||||||||
Sbjct  12720893  TTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCAT  12720854
```

***AtPCP-Bγ* in pET-32a**

NNNNNNNNNNNNNNCNNNNCTANNNTNNTTNNNTTTANCTTNAAGAAGGAGATATACATATGAGCGATAA
AATTATTCNCCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTCGATT
TCTGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAG
GGCAAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGCATCCGTG
GTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGGT
CAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTCTGGTTCTGGCCATATGCACCATCATCATCATCAT
TCTTCTGGTCTGGTGCCACGCGGTTCTGGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACAT
GGACAGCCCAGATCTGGGTACCGACGACGACGACAAGGCCATGGGCGAAAATGGAAAAAGTGTTGAAGC
GAGCAATGCAGCTAAAACACTTTGCATGTCAGTGAATTGCGATAATAAAGACAGAAATCTCACTTGTGCTT
GTTGTTTGGCCAAATCCAAGAATAGGTGTTATAGTAGCAAATCAGAATGTGTTGCTGACTGCAAAGATCTC
GAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTTGGCTGCT
GCCACCGCTGANCAATAACTAGCATAACCCCTTGNNGCCTCTAAACGGGTCTTGAGGGGTTTTTTNGCTGA
AAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCANTTAAGCGCGGCGGGTGT
GGTGNNNNCNCGCAGCGTGACCGCNTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTNCTTCCC
TTCCTTTTCTCGCCACGNTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGNTTCCGA
TTTAGNNGCTTNNNGGCANCTCNACCCCAAAAANNTNNATNNGGGNNGATGGTTCACGTANTGGGCNATC
NCCCTGATANANNNTTTTNCNNCCNTTGACGTTTGNANTCCACGTTCTTTANNANTNNNNTNNNNNNTNCN
AACNGGNNNANNNTCANCNNNNNCGGNNNANTCTTTNNTTNNNNGANTTGCNATTTCGNNATNNNNNAAA
AATGNNCTGANNNNNAATTNACNNNAANTTTANNNNANTNNGCTTACNNNNNNNNNNNNCNTTTTNNGGNN
ANNGNCNNNNNNNCCNANTTNNNNNNTNNNNNNNNNNNNNNNNNNTNNNNCNNNNNNNNNNNNNNNNN
NNNNN

***AtPCP-Bγ* sequencing data alignment**

Arabidopsis thaliana chromosome 2, complete sequence

Sequence ID: ref|NC_003071.7| **Length:** 19698289 **Number of Matches:** 1

Range 1: 7166110 to 7166271 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
300 bits(162)	7e-80	162/162(100%)	0/162(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  543       GAAAATGGAAAAAGTGTTGAAGCGAGCAATGCAGCTAAAACACTTTGCATGTCAGTGAAT  602
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7166271   GAAAATGGAAAAAGTGTTGAAGCGAGCAATGCAGCTAAAACACTTTGCATGTCAGTGAAT  7166212

Query  603       TGCGATAATAAAGACAGAAATCTCACTTGTGCTTGTTGTTTGGCCAAATCCAAGAATAGG  662
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7166211   TGCGATAATAAAGACAGAAATCTCACTTGTGCTTGTTGTTTGGCCAAATCCAAGAATAGG  7166152

Query  663       TGTTATAGTAGCAAATCAGAATGTGTTGCTGACTGCAAAGAT  704
                  ||||||||||||||||||||||||||||||||||||||||||
Sbjct  7166151   TGTTATAGTAGCAAATCAGAATGTGTTGCTGACTGCAAAGAT  7166110
```

***AtPCP-Bδ* in pET-32a**

NNNNNNNNNNNNNNNCNNNCTAGNANNNNTTGTTNAACTTTAAGAAGGAGATATACATATGAGCGATAA
AATTATTCACCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTCGATT
TCTGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAG
GGCAAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGCATCCGTG
GTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGGT
CAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTCTGGTTCTGGCCATATGCACCATCATCATCATCAT
TCTTCTGGTCTGGTGCCACGCGGTTCTGGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACAT
GGACAGCCCAGATCTGGGTACCGACGACGACGACAAGGCCATGGGCGAAAATGGAAAAAGTGTTGAATCG
AACAAGGCAATGAAACCAGTTTGCATGCCAGTGAATTGCAACAATAAAGACAAAAAACTCACTTGTGCTT
GTTGTATCGGGGCCAACCCTAGGAATAGGTGCTACAATAGCAGATCACAATGTACGGCTGACTGCAAACTT
CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG
CTGCCACCGCTGANCAATAACTAGCATAACCCCTTGGNGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTG
AAAGGAGGAACTATATCCGGATTGNCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGT
GGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTANCGCCCGCTCCTTTCGCTTTCNNCCCTTC
CTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAATCGGGGGCTCCCTTTAGGGTTCCNATTTANT
GCTTTACGGCACCTCGACCCCAAAAAACTTGNNTAGGGNGATGGNNCACGTANNGGGCCATCNNCCNNNN
NNNACNGGTTTTCNNCCCNTTGANNTTGGANNNNNCGTTTCTTTTNANAGTGGAANNCTTGNTTCCAAACT
GGGAANACNCTCAACCCNNNNNNCGGTCNANTCCTTTNGATTTNNNAAGNNNTTTTGCNNNNNTCNGNNN
NNNNNNAAAAAANNNNCNNNAATTTANNAAAAATTNACNNCNAATTTTNNCAAANNNNTTNACNCNTNN
NNNTNNGGNNNNNNNNTTNNGNNNNNNNNNNNNNNAACCNNNNNNNNTTNNNTNNNNNAANNNCNNNN
NAANNANNNNNNNCCNNNCNNCNNGNNNNNNCNNNN

***AtPCP-Bδ* sequencing data alignment**

Arabidopsis thaliana chromosome 2, complete sequence

Sequence ID: ref|NC_003071.7| **Length:** 19698289 **Number of Matches:** 1

Range 1: 7157368 to 7157532 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
305 bits(165)	1e-81	165/165(100%)	0/165(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  542      GAAAATGGAAAAAGTGTTGAATCGAACAAGGCAATGAAACCAGTTTGCATGCCAGTGAAT  601
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7157532  GAAAATGGAAAAAGTGTTGAATCGAACAAGGCAATGAAACCAGTTTGCATGCCAGTGAAT  7157473

Query  602      TGCAACAATAAAGACAAAAAACTCACTTGTGCTTGTTGTATCGGGGCCAACCCTAGGAAT  661
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7157472  TGCAACAATAAAGACAAAAAACTCACTTGTGCTTGTTGTATCGGGGCCAACCCTAGGAAT  7157413

Query  662      AGGTGCTACAATAGCAGATCACAATGTACGGCTGACTGCAAACTT  706
                |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7157412  AGGTGCTACAATAGCAGATCACAATGTACGGCTGACTGCAAACTT  7157368
```

***AtPCP-Bβ* 2 in pJET1.2**

NNNNNNNNNTCGGATGGCTCGAGTTTTTCNNCNGATGTCATCATCACATTTTGTTATCCTTTGCTTGATCGT
AGTTTCTCTCGTTCCTCATGGATCTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAGCTGCAAACG
CACAAGCTGTTGACGCGACCAAATCAGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATC
AGTGTGCCATCTTGGCAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCA
GGTGTTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATATCTTTCTAGAAGATCTCCT
ACAATATTCTCAGCTGCCATGGAAAATCGATGTTCTTCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAA
AACTTATATTAAGAACTATGCTAACCACCTCATCAGGAACCGTTGTAGGTGGCGTGGGTTTTCTTGGCAATC
GACTCTCATGAAAACTACGAGCTAAATATTCAATATGTTCCTCTTGACCAACTTTATTCTGCATTTTTTTTGA
ACGAGGTTTAGAGCAAGCTTCAGGAAACTGAGACAGGAATTTTATTAAAAATTTAAATTTTGAAGAAAGTT
CAGGGTTAATAGCATCCATTTTTTGCTTTGCAAGTTCCTCAGCATTCTTAACAAAAGACGTCTCTTTTGACA
TGTTTAAAGTTTAAACCTCCTGTGTGAAATTATTATCCGCTCATAATTCCACACATTATACGAGCCGGAAGC
ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCAAT
TGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG
TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCNGGGGATAACGCNNNAAGANATGTGAG
CAAAAGNCAGCAAAGGCCAGGANCGTAAAAAGCCGCGTTGCTGGCGTTTTCCATAGGCTCGCCCCCTGAC
GAGCATCANAAAATCGACGCTCAGTCAGAGTGNNAANCCGANNGNCTANNAGATNCNGNGTTNCCCCNNN
NCTNCCTCNNNNNNNNCNNNCNACNNNNGNTACGNNNCNNNCGCTTNNNCNNGNNNNNNNTNNNNNNNN
NANNNNNNGNNNNNNNTNNNGNAGNCNNNNNNNNNNGNNNNNNNNNNNNNNNNNNNNNNANNNNNNN
NNNGNANNTNCNNGNNNNCNNNNNANNNNNNNNNNNCCNNN

AtPCP-Bβ 2 sequencing data alignment

Range 1: 12720854 to 12721022 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
313 bits(169)	8e-84	169/169(100%)	0/169(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  169       AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  228
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721022  AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  12720963

Query  229       CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  288
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720962  CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  12720903

Query  289       TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCAT  337
                  ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720902  TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCAT  12720854
```

Range 2: 12721810 to 12721870 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
113 bits(61)	9e-24	61/61(100%)	0/61(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  35        ATGTCATCATCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTCATGGA  94
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721870  ATGTCATCATCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTCATGGA  12721811

Query  95        T  95
                  |
Sbjct  12721810  T  12721810
```

Range 3: 12721624 to 12721662 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
73.1 bits(39)	2e-11	39/39(100%)	0/39(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  96        CTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAG  134
                 |||||||||||||||||||||||||||||||||||||||
Sbjct  12721662  CTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAG  12721624
```

Range 4: 12721332 to 12721369 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
71.3 bits(38)	6e-11	38/38(100%)	0/38(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  133       AGCTGCAAACGCACAAGCTGTTGACGCGACCAAATCAG  170
                  ||||||||||||||||||||||||||||||||||||||
Sbjct  12721369  AGCTGCAAACGCACAAGCTGTTGACGCGACCAAATCAG  12721332
```

AtPCP-Bβ 3 in pJET1.2

NNNNNNNNNTCNGNTGGCTCGAGTTTTTCNNCNNATTTCGGATTTCAACCATCAAACATAAACAAAAAAA
AAACCAAGAAAACAACAATGTCATCATCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTC
ATGGATCTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAGCTGCAAGCGCACAAGCTGTTGACGC
GACCAAATCAGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGGCA
AATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTGTTATCGTAGTTTA
TATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATATCTTTCTAGAAGATCTCCTACAATATTCTCAGCTG
CCATGGAAAATCGATGTTCTTCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAAAACTTATATTAAGAA
CTATGCTAACCACCTCATCAGGAACCGTTGTAGGTGGCGTGGGTTTTCTTGGCAATCGACTCTCATGAAAAC
TACGAGCTAAATATTCAATATGTTCCTCTTGACCAACTTTATTCTGCATTTTTTTTGAACGAGGTTTAGAGCA
AGCTTCAGGAAACTGAGACAGGAATTTTATTAAAAATTTAAATTTTGAAGAAAGTTCAGGGTTAATAGCAT
CCATTTTTTGCTTTGCAAGTTCCTCAGCATTCTTAACAAAAGACGTCTCTTTTGACATGTTTAAAGTTTAAAC
CTCCTGTGTGAAATTATTATCCGCTCATAATTCCACACATTATACGAGCCGGAAGCATAAAGTGTAAAGCC
TGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCAATTGCTTTCCAGTCGGGA
AACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGANGCGGTTTGCGTATTGGGCGCTC
TTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGNTGNGCGAGCGGTATCAGCTCACTCAAAG
GCGGTAATACGGTNATCANAGATCAGGGNNACGCANGAAANACNTGTNANCAAANGCCAGNAANNNGCN
GNAACCGNNAAAANGCNNNNTGCTGNNGNNTTNNNTGANNNGCCCCCNTGNNNNNNTNAANNAANNNNC
NNNNAANNNNNNANGNNGCNANNNNNANNNNNNAAANNNNNGNCNGNCNNTNCNNNNNAANNNNNNNN
NNNNCTNNNNNNNN

***AtPCP-Bβ* 3 sequencing data alignment**

Range 1: 12720854 to 12721022 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
313 bits(169)	8e-84	169/169(100%)	0/169(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  222       AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  281
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721022  AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  12720963

Query  282       CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  341
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720962  CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  12720903

Query  342       TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCAT  390
                  ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720902  TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCAT  12720854
```

Range 2: 12721810 to 12721921 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
202 bits(109)	2e-50	111/112(99%)	0/112(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  37        TTCGGATTTCAACCATCAAACATAAACaaaaaaaaaaCCAAGAAAACAACAATGTCATCA  96
                 |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct  12721921  TTCGGATTTCAACCATCAAACATAAACAAAAAAAAAACCAAAAAAACAACAATGTCATCA  12721862

Query  97        TCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTCATGGAT  148
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721861  TCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTCATGGAT  12721810
```

Range 3: 12721624 to 12721662 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
73.1 bits(39)	1e-11	39/39(100%)	0/39(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  149       CTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAG  187
                  |||||||||||||||||||||||||||||||||||||||
Sbjct  12721662  CTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAG  12721624
```

Range 4: 12721332 to 12721369 GenBank Graphics ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
65.8 bits(35)	2e-09	37/38(97%)	0/38(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  186       AGCTGCAAGCGCACAAGCTGTTGACGCGACCAAATCAG  223
                  |||||||| ||||||||||||||||||||||||||||||
Sbjct  12721369  AGCTGCAAACGCACAAGCTGTTGACGCGACCAAATCAG  12721332
```

***AtPCP-Bβ* 4 in pJET1.2**

NNNNNNNNTNNGNTGGCTCGAGTTTTTCNGCAAGATTTGTAGCTGTATTGTTTGTATAAGGTCGTTAATTTA
ATGAACAAGCCGGTTGCAGACAGCCACACACTTATATAAACTACGATAACACCTGCTACGATCATTGAAAC
AGCAATAACACACTTCTTCTCTGTGTTTTGGACATTTGCCAAGATGGCACACTGATGAACCTAACTTGTTCG
CGTCAAAACCTTGATCTTTCTAGAAGATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGATGTTCTTCT
TTTATTCTCTCAAGATTTTCAGGCTGTATATTAAAACTTATATTAAGAACTATGCTAACCACCTCATCAGGA
ACCGTTGTAGGTGGCGTGGGTTTTCTTGGCAATCGACTCTCATGAAAACTACGAGCTAAATATTCAATATGT
TCCTCTTGACCAACTTTATTCTGCATTTTTTTTGAACGAGGTTTAGAGCAAGCTTCAGGAAACTGAGACAGG
AATTTTATTAAAAATTTAAATTTTGAAGAAAGTTCAGGGTTAATAGCATCCATTTTTTGCTTTGCAAGTTCC
TCAGCATTCTTAACAAAAGACGTCTCTTTTGACATGTTTAAAGTTTAAACCTCCTGTGTGAAATTATTATCC
GCTCATAATTCCACACATTATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGACTNANTNNTGANTT
GACTCANATTAATTGCGTTACNNTNNCTGCNANNNNNNTNNNNNGCTGNAATGCTGNCNNGNCNNCTGAN
NTCANGAANANGCCAACGNGCGTNGCGAGNTGGTTTGCGACTCGGGTGCCCTTACGCTTCCTGGATCACTG
ANNCACTGAGACAACTCGTTNCGTNCNTTGNGNNGNTNANCTNNNTNNACGCTNNNNNCNGTTNTCACAG
AATCGTGACGTAACANNCAAGCAAAACTGANCNCAGGNANNNAANNNNNCAACCGTAANNGNNGCGTGC
TGNNNNANNNNNNNNCTNNNCNCNAGAATACNTGCNANGNNNTCNNTCNNNCAGANNAGCCAACNNNNN
AGNCTNNNNATGACTGNCNTNNCNCNNNNGNNNCGCGAGNTCANNCTGATCCNCNNGCNNCTANNNNTAC
NNGTNAGCTNNGNCNNNNNNNNNNNNGNGCTNNNNANANATGANNNNNNNATCNNANNNNNNNAGTNN

NTNNNNNANNNNNCNNNNNNNNNNACNCNNNNNNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNTNN
NNNGNNNNNNNNCANNNNNNNNNNNNNNNNNNNNNNNNGNANNGANANNNNTNCGNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNANNN

AtPCP-Bβ 4 sequencing data alignment

Range 1: 12720818 to 12721009 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
355 bits(192)	1e-96	192/192(100%)	0/192(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  1166      CAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGGCAAATGTCCAAAA  1225
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721009  CAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGGCAAATGTCCAAAA  12720950

Query  1226      CACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTGTTATCGTAGTTTA  1285
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720949  CACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTGTTATCGTAGTTTA  12720890

Query  1286      TATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGACCTTATACAAACA  1345
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720889  TATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGACCTTATACAAACA  12720830

Query  1346      ATACAGCTACAA  1357
                 ||||||||||||
Sbjct  12720829  ATACAGCTACAA  12720818
```

AtPCP-Bβ 5 in pJET1.2

NNNNNNNNTCGGATGGCTCGAGTTTTTCNNCANATGTCATCATCACATTTTGTTATCCTTTGCTTGATCGTA
GTTTCTCTCGTTCCTCATGGATCTGGAGACAAGACACACAGTTTTGACGCCAACAAGGCAGCTGCAAACGC
ACAAGCTGTTGACGCGACCAAATCAGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCA
GTGTGCCATCTTGGCAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAG
GCGTTATCGTAGTTTATATAAGTGCGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGACCTTATACA
AACAATACAGCATCTTTCTAGAAGATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGATGTTCTTCTTT
TATTCTCTCAAGATTTTCAGGCTGTATATTAAAACTTATATTAAGAACTATGCTAACCACCTCATCAGGAAC
CGTTGTAGGTGGCGTGGGTTTTCTTGGCAATCGACTCTCATGAAAACTACGAGCTAAATACTGATGAANNC
CCTAATGATTTTNATTATAATNNNTTAGTTTNGGNAGANNCNCATCNTGNCCTATNATCAAATGGNTTCGG
CAATTATCNATAANNTTACATTTNAANGTGCGAACTCGATNTTTTACACNACTCTCTTTACCNATTCTGCCC
CCAATTACACTTAAAACNACNCAACAGCTTAACGTNNGCTTGCCNCGCCTTACTTGACTGAAAACTCTCAC
TCTTACCGAACTTGGCCGTAACCTGNCNNNAAAGCGAGAACNAAACNTAACATCAAACGAATCGACCGAT
TGNNNNGTAATCGTCACCTCCACANAGAGCGACTCTCTGTATACCGTTGGCATGCTANCTTTATCTGTTCGG
GCAATACNATGCCCATTGTACTTGTTGACTGGTCTGATATCCGTGAGCAAAACGGCTTATGGNATTGCGAG
CTTCAGTCGCACTACACNNNCGTTCTGTNCTCTTTATGAGANAGCNTTCCCGCTTTCNNANCAATGNTCNAN
AAANNTCATGACCAATTTCTAGCCGACCTTGCGAGCATNNNCCCAGNANNNCCNCACCGCTCANNGNNNN
NGATGNNGNNTTNNNNNNCNNGGNNNNNNCCNTNAGAANCNNNNNNGTACTNNTNANNNNAGNANNNGN
AAANNCANNNNGNNNACNNNNNCGGANANTGNNNNNNTNNNCNCTNNNNNNNNNNNTNNNNNNCNNTN
NNNANTTNNNNNNNNAAGNNNNNNAANNANCNNNNNNNNNNNNNNNNNNNNANNNNNNANNNCNNANA
NNNNNNNNNNNNNNNNNNCNCNNNCNNNNNNNACNNNNNNNNNNNNANTNNNNNTNACNNNNNNNNN

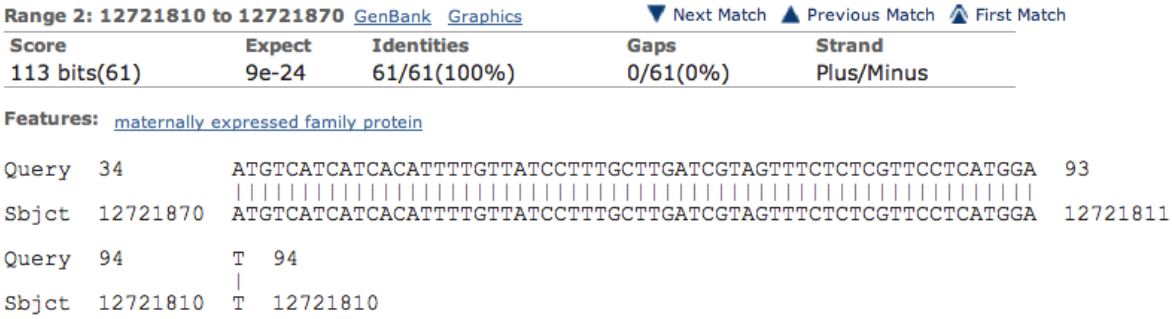
AtPCP-Bβ 5 sequencing data alignment

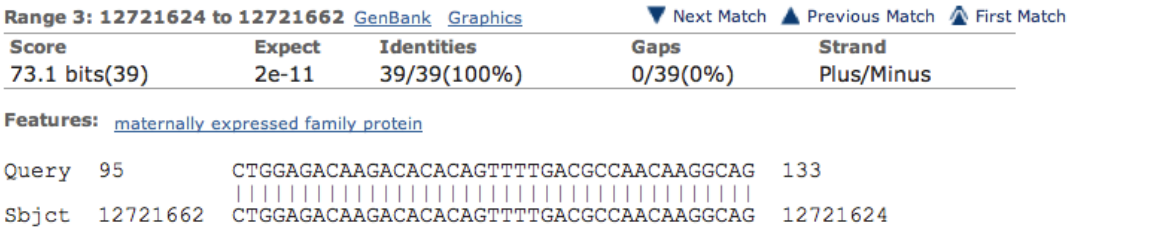
Range 1: 12720823 to 12721022 GenBank Graphics ▼ Next Match ▲ Previous Match

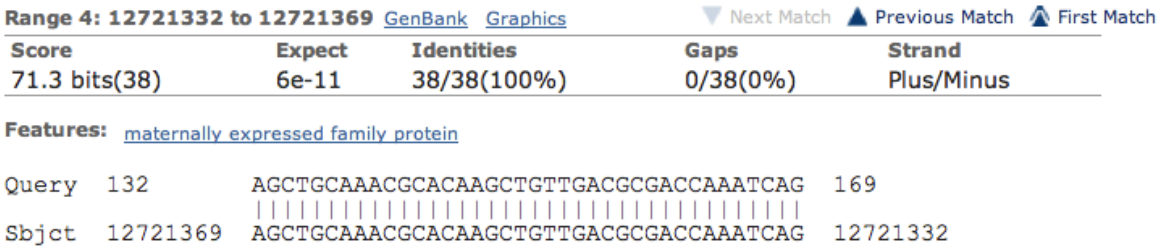
Score	Expect	Identities	Gaps	Strand
359 bits(194)	1e-97	198/200(99%)	0/200(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  168       AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  227
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721022  AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  12720963

Query  228       CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGCG  287
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  12720962  CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  12720903

Query  288       TTATCGTAGTTTATATAAGTGCGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGA  347
                 |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  12720902  TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGA  12720843

Query  348       CCTTATACAAACAATACAGC  367
                 ||||||||||||||||||||
Sbjct  12720842  CCTTATACAAACAATACAGC  12720823
```







***AtPCP-Bβ* 6 in pJET1.2**

NNNNNNNTCGGATGGCTCGAGTTNNNNNCAGATTTGTAGCTGTATTGTTTGTATAAGGTCGTTAATTTAAT
GAACAAGCCGGTTGCAGACAGCCACACACTTATATAAACTACGATAACACCTGCTACGATCATTGAAACAG
CAATAACACACTTCTTCTCTGTGTTTTGGACATTTGCCAAGATGGCACACTGATGAACCTAACTTGTTCGCG
TCAAAACCTTGCCCGTTTGCATCTGATTTGGTCGCGCCAACAGCTTGTGCGTTTGCAGCTTGTTGGCGTCAA
AACTGTGTGTCTTGTCTCCAGATCCATGAGGAACGAGAGAAACTACGATCAAGCAAAGGATAACAAAATG
TGATGATGACATTGTTGTTTTTTTGGTTTTTTTTTTTGTTTATGTTTGATGGTTGAAATCCGAATCTTTCTAGAA
GATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGATGTTCTTCTTTTATTCTCTCAAGATTTTCAGGCTG
TATATTAAAACTTATATTAAGAACTATGCTAACCACCTCATCAGGAACCGTTGTAGGTGGCGTGGGTTTTCT
TGGCAATCGACTCTCATGAAAACTACGAGCTAAATATTCAATATGTTCCTCTTGACCAACTTTATTCTGCAT
TTTTTTTGAACGAGGTTTAGAGCAAGCTTCAGGAAACTGAGACAGGAATTTTATTAAAAATTTAAATTTTGA
AGAAAGTTCAGGGTTAATAGCATCCATTTTTTGCTTTGCAAGTTCCTCAGCATTCTTAACAAAAGACGTCTC
TTTTGACATGTTTAAAGTTTNNCCTCCTGTGTGAAATTATTATCCGCTCATAATTCCACACATTATACGAGC
CGGAAGCATAAAGTGTAAAGCCNGGANGANNCCNCGANTGATCTTNATCANAATCATTAAGTTNNGGTAG
ATACNCATCTNGTCNNATGATCAAATGGTGTCGNCGNNAATCAATAACNGACTACAAAAGTGCNAACTCG
ATATTNTACGCGACTCGNNNTACCATTCNGNCCCGAATACACTNNAACGACTCACANCTTACGTGNNTGCC
CGNCTTACTGACTGNANACTCTNNCTCTNCNNACTGNNCNNNNNTGNCACCNGANCGAGACNNACNAANG
CNACNATGNNGATGNAGNATNNNCNCTCNGNANNNNNCTNNNNTCNNCCNNNGNANGNTAGCTNANNNG
TCNGNNANNNANNNNNNNNNNNNNGTNANNGNNNTGANNNCCNNNGCNNACCNCNNAGNNTNTNNNNNNT
NNNNGNTNNACNNNNNNNNNCNNNTNNNNNNAAANNNNNNNNNNNNNNNNNNNNNNNNNNNACNNT
NNNNCNANNNNNNNNNTNNNNNNNNANCNNNNNNNNNNNNNNNNNCNNNNTTNNN

***AtPCP-Bβ* 6 sequencing data alignment**

Range 1: 12720818 to 12721022 GenBank Graphics — Next Match / Previous Match

Score	Expect	Identities	Gaps	Strand
379 bits(205)	8e-104	205/205(100%)	0/205(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  1168      AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  1227
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721022  AGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGGTTCATCAGTGTGCCATCTTGG  12720963

Query  1228      CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  1287
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720962  CAAATGTCCAAAACACAGAGAAGAAGTGTGTTATTGCTGTTTCAATGATCGTAGCAGGTG  12720903

Query  1288      TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGA  1347
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12720902  TTATCGTAGTTTATATAAGTGTGTGGCTGTCTGCAACCGGCTTGTTCATTAAATTAACGA  12720843

Query  1348      CCTTATACAAACAATACAGCTACAA  1372
                 |||||||||||||||||||||||||
Sbjct  12720842  CCTTATACAAACAATACAGCTACAA  12720818
```

Range 2: 12721810 to 12721921 GenBank Graphics — Next Match / Previous Match / First Match

Score	Expect	Identities	Gaps	Strand
207 bits(112)	4e-52	112/112(100%)	0/112(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  987       TTCGGATTTCAACCATCAAACATAAACaaaaaaaaaaccaaaaaaaCAACAATGTCATCA  1046
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721921  TTCGGATTTCAACCATCAAACATAAACAAAAAAAAAACCAAAAAAACAACAATGTCATCA  12721862

Query  1047      TCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTCATGGAT  1098
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12721861  TCACATTTTGTTATCCTTTGCTTGATCGTAGTTTCTCTCGTTCCTCATGGAT  12721810
```

Range 3: 12721628 to 12721662 GenBank Graphics — Next Match / Previous Match / First Match

Score	Expect	Identities	Gaps	Strand
65.8 bits(35)	3e-09	35/35(100%)	0/35(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  1099      CTGGAGACAAGACACACAGTTTTGACGCCAACAAG  1133
                 |||||||||||||||||||||||||||||||||||
Sbjct  12721662  CTGGAGACAAGACACACAGTTTTGACGCCAACAAG  12721628
```

Range 1: 12721332 to 12721369 GenBank Graphics — Next Match / Previous Match

Score	Expect	Identities	Gaps	Strand
67.9 bits(34)	1e-11	37/38(97%)	0/38(0%)	Plus/Minus

Features: maternally expressed family protein

```
Query  2         AGCTGCAAACGCACAAGCTGTTGGCGCGACCAAATCAG  39
                 ||||||||||||||||||||||| ||||||||||||||
Sbjct  12721369  AGCTGCAAACGCACAAGCTGTTGACGCGACCAAATCAG  12721332
```

2.5 Amino acid sequences translated from *At2g29790* alternative splicing transcripts cloning products

ORF 1

```
1   ttcggatttcaaccatcaaacataaacaaaaaaaaaaccaaaaaaacaacaatgtcatca
1    F  G  F  Q  P  S  N  I  N  K  K  K  T  K  K  T  T  M  S  S

61  tcacattttgttatcctttgcttgatcgtagtttctctcgttcctcatggatCTGGAGAC
21   S  H  F  V  I  L  C  L  I  V  V  S  L  V  P  H  G  S  G  D

121 AAGACACACAGTTTTGACGCCAACAAGgcagCTGCAAACGCACAAGCTGTTGACGCGACC
41   K  T  H  S  F  D  A  N  K  A  A  A  N  A  Q  A  V  D  A  T

181 AAATCAGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGgttcatcagtgtgccat
61   K  S  D  A  N  G  Q  G  F  D  A  N  K  L  G  S  S  V  C  H

241 cttggcaaatgtccaaaacacagagaagaagtgtgttattgctgtttcaatgatcgtagc
81   L  G  K  C  P  K  H  R  E  E  V  C  Y  C  C  F  N  D  R  S

301 aggtgttatcgtagtttatataagtgtgtggctgtctgcaaccggcttgttcattaaatt
101  R  C  Y  R  S  L  Y  K  C  V  A  V  C  N  R  L  V  H  *  I

361 aacgaccttatacaaacaatacag
121  N  D  L  I  Q  T  I  Q
```

ORF 2

```
1   tcggatttcaaccatcaaacataaacaaaaaaaaaaccaaaaaaacaacaatgtcatcat
1    S  D  F  N  H  Q  T  *  T  K  K  K  P  K  K  Q  Q  C  H  H

61  cacattttgttatcctttgcttgatcgtagtttctctcgttcctcatggatCTGGAGACA
21   H  I  L  L  S  F  A  *  S  *  F  L  S  F  L  M  D  L  E  T

121 AGACACACAGTTTTGACGCCAACAAGgcagCTGCAAACGCACAAGCTGTTGACGCGACCA
41   R  H  T  V  L  T  P  T  R  Q  L  Q  T  H  K  L  L  T  R  P

181 AATCAGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGgttcatcagtgtgccatc
61   N  Q  M  Q  T  G  K  V  L  T  R  T  S  *  V  H  Q  C  A  I

241 ttggcaaatgtccaaaacacagagaagaagtgtgttattgctgtttcaatgatcgtagca
81   L  A  N  V  Q  N  T  E  K  K  C  V  I  A  V  S  M  I  V  A

301 ggtgttatcgtagtttatataagtgtgtggctgtctgcaaccggcttgttcattaaatta
101  G  V  I  V  V  Y  I  S  V  W  L  S  A  T  G  L  F  I  K  L

361 acgaccttatacaaacaatacagc
121  T  T  L  Y  K  Q  Y  S
```

ORF 3

```
1   cggatttcaaccatcaaacataaacaaaaaaaaaaccaaaaaaacaacaatgtcatcatc
1    R  I  S  T  I  K  H  K  Q  K  K  N  Q  K  N  N  N  V  I  I

61  acattttgttatcctttgcttgatcgtagtttctctcgttcctcatggatCTGGAGACAA
21   T  F  C  Y  P  L  L  D  R  S  F  S  R  S  S  W  I  W  R  Q

121 GACACACAGTTTTGACGCCAACAAGgcagCTGCAAACGCACAAGCTGTTGACGCGACCAA
41   D  T  Q  F  *  R  Q  Q  G  S  C  K  R  T  S  C  *  R  D  Q
```

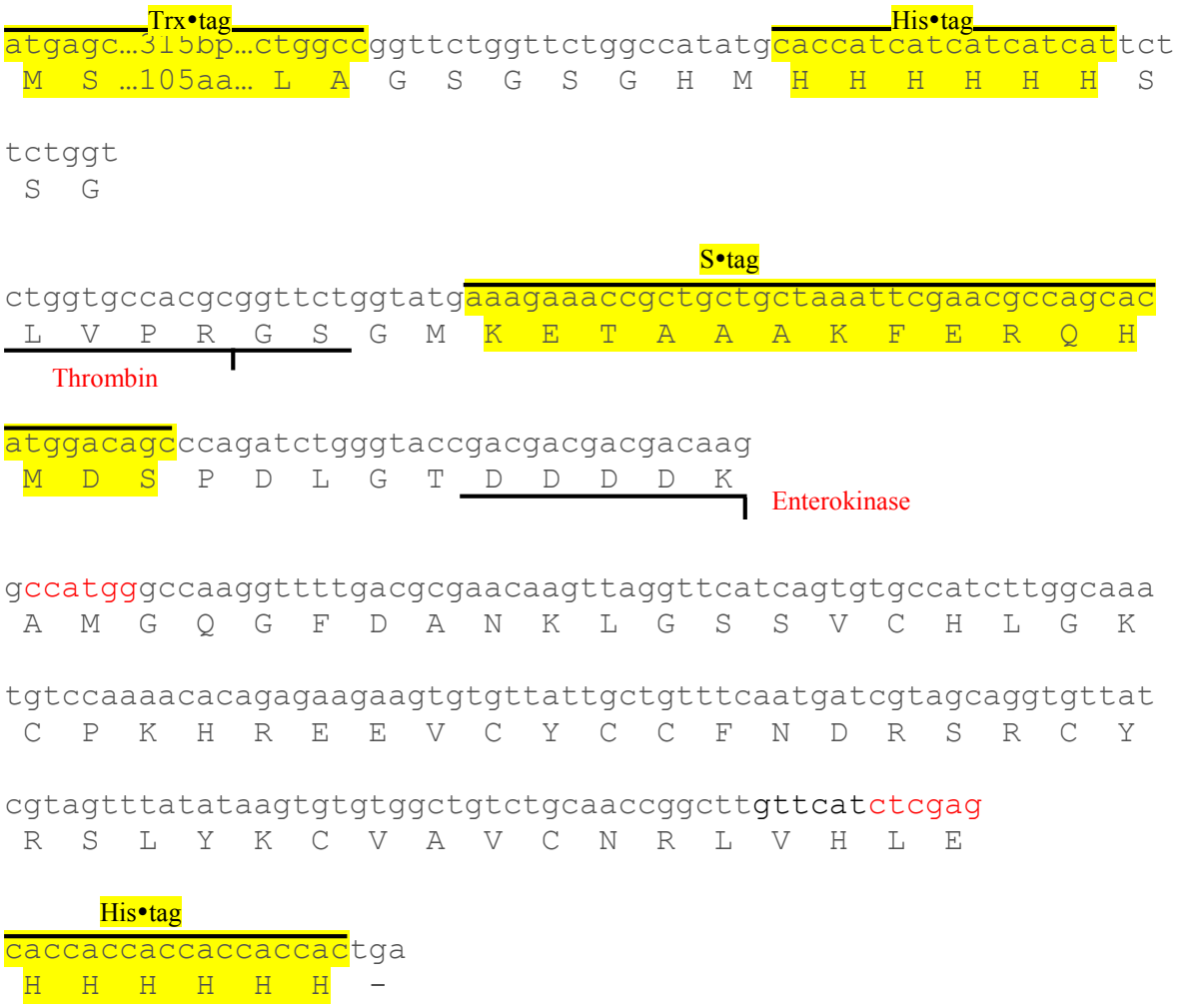
```
181 ATCAGATGCAAACGGGCAAGGTTTTGACGCGAACAAGTTAGgttcatcagtgtgccatct
61   I  R  C  K  R  A  R  F  *  R  E  Q  V  R  F  I  S  V  P  S

241 tggcaaatgtccaaaacacagagaagaagtgtgttattgctgtttcaatgatcgtagcag
81   W  Q  M  S  K  T  Q  R  R  S  V  L  L  L  F  Q  *  S  *  Q

301 gtgttatcgtagtttatataagtgtgtggctgtctgcaaccggcttgttcattaaattaa
101  V  L  S  *  F  I  *  V  C  G  C  L  Q  P  A  C  S  L  N  *

361 cgaccttatacaaacaataca
121  R  P  Y  T  N  N  T
```

2.6 Amino acid sequences of expressed AtPCP-Bs fusion proteins

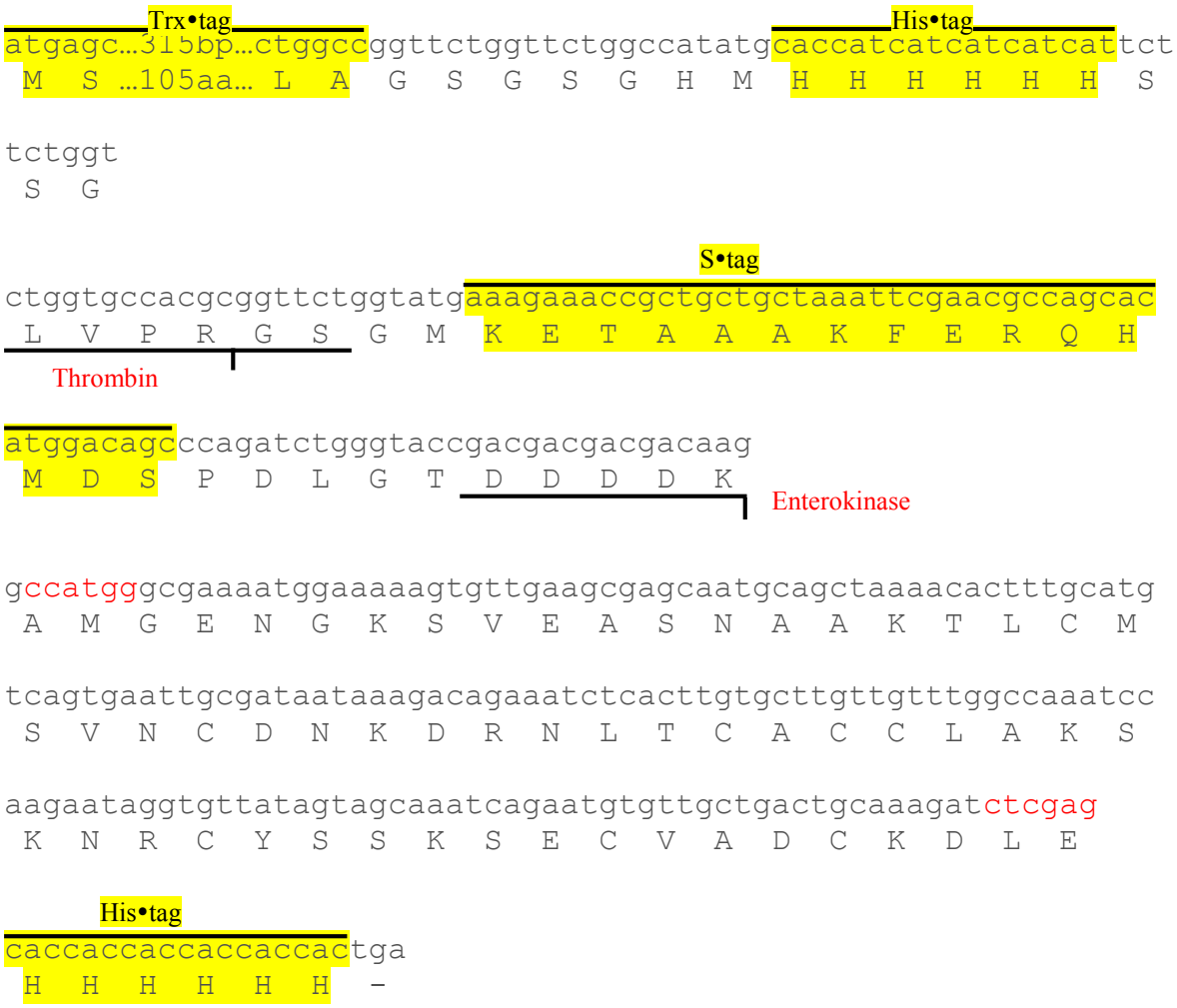
AtPCP-Bα (17.78+6.73=24.51kD)

```
          Trx•tag                                                  His•tag
atgagc…315bp…ctggccggttctggttctggccatatgcaccatcatcatcatcattct
 M   S …105aa… L   A   G   S   G   S   G   H   M   H   H   H   H   H   H   S

tctggt
 S   G

                                                    S•tag
ctggtgccacgcggttctggtatgaaagaaaccgctgctgctaaattcgaacgccagcac
 L   V   P   R   G   S   G   M   K   E   T   A   A   A   K   F   E   R   Q   H
                  |
     Thrombin

atggacagccagatctgggtaccgacgacgacgacaag
 M   D   S   P   D   L   G   T   D   D   D   D   K
                                                  |
                                                  Enterokinase

gccatgggctctgaggcggagctgcaattagatccatcaatgtgtctgcgtgtcgaatgt
 A   M   G   S   E   A   E   L   Q   L   D   P   S   M   C   L   R   V   E   C

gcaaaacacagaaatcaaaaatggtgtttttgctgtgccggactaccgaggacctgtttt
 A   K   H   R   N   Q   K   W   C   F   C   C   A   G   L   P   R   T   C   F

cttgataaacgaggctgtacgtctgtctgcaagcgcgaatctccttcaatggctctcgag
 L   D   K   R   G   C   T   S   V   C   K   R   E   S   P   S   M   A   L   E

        His•tag
caccaccaccaccaccactga
 H   H   H   H   H   H   -
```

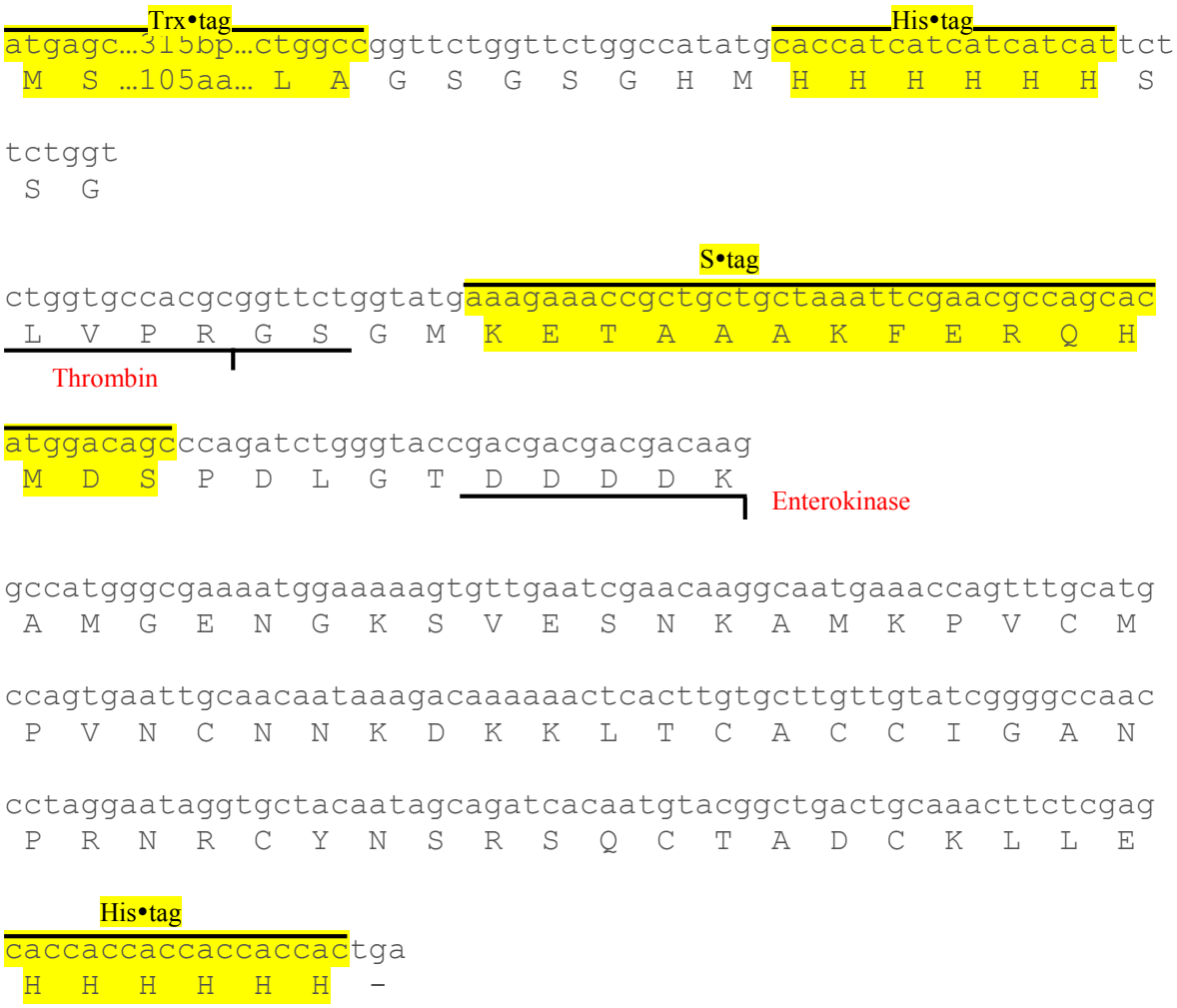
AtPCP-Bβ (17.78+6.50=24.28kD)



AtPCP-Bγ (17.78+6.31=24.09kD)



AtPCP-Bδ (17.78+6.54=24.32kD)



2.7 *At2g29790* and *AtPCP-B*, Chromosome 2 complement strand

Position

12722800 tgcatagagaagatatatgtcatttaatttgtaacattttctctagactcactcatttttgcaaagtatttatttgtaacagtttccaacaatttttgc

aaagtatttatttgtcatgttaaacaacaacaatatcactaaattactaaggaatacggaagaactaagactagtaaaaacaattaatttaacagttttt

12722600 tacggatcaccataagagacatgcactaataattacaaacaacttatctataaaaacaaactgtatgcaaggagttcaggttttcaaccatcaaataacc

caaaaaaaaaaaaaaaacagaaaaccaaaaaccaaaaaaaaaaaaaaaaagtttctgatttttagaagaATGTCACCATCACATTTTGCTATCCTTTTCAT

12722400 CATCGTAATTTCTCTGGTTCCTCTACATGGATgtacgtctatgagtttatgcatatatatacataatcgctttgtcattgtttacaatcagtcatcacca

tattgttatatataattctccagatcaattattaacttctatttttttttccttatagCTGGAAACACACAACATCATAACGCGAACAAGCCAGgtgtgga

12722200 ttttaactctatctaataacaaacaaataaacgaaaaaacaaaaaaccgaatatcacaaacaatgtcatcgtcaaattttggtatcatttgcatcatact

gattttctcattcctctacaagaatgtatttatatgaaatataatcgttttcatcttttacaattatagtcataaccattaagaaatcatattgttatta

12722000 ataaaaattcattttctttaactatttttttttttcttatagCTCCAAACGCACAAGCTGTTGACGCAAGCAAATCAGgttcggatttcaaccatcaaac

ataaacaaaaaaaaaaccaaaaaaacaacaatgtcatcatcacattttgttatcctttgcttgatcgtagtttctctcgttcctcatggatgtatgttca

12721800 taaatttatgcatatatacaaataattgccttttcattgtttacaatcagtcatcaccattataatcatatacttataaatatttatttttatatttttt

tggtaaaaaataaatattaacttatatttttttaatagCTGGAGACAAGACACACAGTTTTGACGCCAACAAGgcaggtaattcggatttcaactaccaa

12721600 ataaccaaacaaatatttgtgttttaaaaaagaaagaatatcacaacaacgtcatcatcagattttgttataatttgcatcatcgtaattttctcgttcc

tctacaagaatttatgaaattaaaggctttttcttcgtttacattcatcaccattgtaaaatatgttgttagaaataaaattattcgctttttttaactct

12721400 tcattttcttaattctaatttttttttcttatagCTGCAAACGCACAAGCTGTTGACGCGACCAAATCAGgttcggaatatttcaaccatcatatagcaaa

caaagaattaacaaaaagaaaaaaaaatacgaatataaaaagaatatcatcatcacattttgttatgatcctttaatttgcatcatcgtaatttctcttg

12721200 ttcctctacaaaaatatatgtttttgaaattatgcgtatatatataaataaggaatcgcttattcatcgtttacaatcatcaccattgtaatcgtattgt

tataaataaaattcttctttttaatttttttttttaatttttttttttttttttttgaacttctaatctttttcttatagATGCAAACGGGCAAGGTTTT

12721000 GACGCGAACAAGTTAGgttcatcagtgtgccatcttggcaaatgtccaaaacacagagaagaagtgtgttattgctgtttcaatgatcgtagcaggtgtt

atcgtagtttatataagtgtgtggctgtctgcaaccggcttgttcattaaattaacgaccttatacaaacaatacagctacaagtttacctaaagtaaac

12720800 atatatatacaatataattatgcttatcgtaattttgtttagtgccattgaaagtttaagtaattttatttttctattaagttttcttaaaaaatattaa

tttattttgaataatatgttgtgtaataagaatttataaattaattagttaatgatgattttgtaccaactaatgatagatcatacttgtatagttgtat

12720600 tatagtaatatttaaaagtaaattttaatatttaggataacaactgaacatatatttttttaacagttgctttcgctaatatctcgttgtaattaagaaaa

gataaaagattgccttgggactttgacatgtgtttttgggtcttcaaccacattctacaggttaaatctaaaaacactaa

ATG = Translational Start/Stop

ATGC = Exon

atgc = Annotation on other strand Hide

atgc = UTR

atgc = Intron

Boxed text = AtPCP-Bβ exon

2.8 Statistically validated hits from mass spectrometry analysis of proteins obtained from the Arabidopsis stigmatic microsomal membrane fraction.

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q39101	Ferritin-1	89.74	6.27	1	2	2	27	5.874E6	255	28.2	6.11
P23321	Oxygen-evolving enhancer protein 1-1	47.72	35.54	2	6	6	16	1.446E6	332	35.1	5.66
P51407	60S acidic ribosomal protein	46.51	41.74	5	3	4	16	1.405E6	115	11.4	4.61
Q8LCW9	60S acidic ribosomal protein	43.83	34.82	1	1	1	9	2.218E6	112	11.2	4.36
Q9SLF7	60S acidic ribosomal protein	43.53	41.74	3	3	4	23	1.843E6	115	11.4	4.68
Q93VG5	40S ribosomal protein S8-1	40.92	33.78	2	6	6	17	9.232E5	222	25.0	10.32
P41127	60S ribosomal protein L13-1	24.33	12.62	3	2	2	6	6.163E5	206	23.8	11.02
P50883	60S ribosomal protein L12-1	23.03	31.33	3	3	3	10	1.270E6	166	17.9	8.97
Q8H135	ATP synthase subunit beta (Fragment)	22.80	21.52	5	6	6	13	4.103E5	446	48.2	5.63
B9DI38	AT1G05190 protein (Fragment)	19.31	11.00	2	1	1	5	4.403E5	200	22.2	9.80
P53492	Actin-7	19.07	7.69	9	1	2	8	1.146E6	377	41.7	5.49
O23095	60S acidic ribosomal protein P1-2	18.78	36.28	1	2	2	4	2.781E5	113	11.3	4.32
A6XI99	Ubiquitin (Fragment)	17.68	23.29	2	1	1	5	4.353E5	219	24.6	6.83
Q93ZL9	AT3g18780/MVE11_16	16.00	8.48	9	1	2	6	1.252E5	342	38.6	6.47
P51430	40S ribosomal protein S6-2	15.22	24.50	4	4	4	6	1.120E5	249	28.1	10.83
Q2V3X4	60S ribosomal protein L4-1	14.85	13.33	4	1	3	7	4.184E5	405	44.5	10.37
P38418	Lipoxygenase 2	14.32	5.36	1	2	2	4	4.107E5	896	102.0	5.62
Q06327	Linoleate 9S-lipoxygenase 1	12.90	2.10	1	1	1	4	2.086E5	859	98.0	5.52
Q8LC58	Photosystem I reaction center subunit IV B	12.69	30.07	2	1	2	5	1.069E5	143	15.0	9.95
F4KDU5	60S ribosomal protein L4-2	12.69	14.04	4	1	3	5	6.328E5	406	44.6	10.35
P29402	Calnexin homolog 1	12.10	6.42	3	2	2	5	2.468E5	530	60.4	4.91
Q9SSB8	Cytochrome c oxidase subunit 5b-2	11.34	23.39	1	1	2	6	2.742E5	171	18.6	5.73
A8MRD7	Cytochrome c oxidase subunit 5b-1	11.34	21.14	2	1	2	5	2.605E5	175	19.3	5.02
Q9FNP8	40S ribosomal protein S19-3	11.21	10.49	3	2	2	5	2.438E5	143	15.7	10.21
Q42218	Ferritin 2 (Fragment)	11.00	12.90	2	1	1	3	6.103E5	124	13.3	10.29
Q9LR33	60S ribosomal protein L27a-2	10.80	12.33	1	1	1	4	5.967E5	146	16.3	10.51
Q9SUT2	Peroxidase 39	10.29	6.75	1	1	1	4	1.705E5	326	35.6	6.98
Q8LEQ0	60S acidic ribosomal protein P1-3	10.18	35.40	1	1	1	2	8.423E5	113	11.2	4.36
Q42589	Non-specific lipid-transfer protein 1	10.05	37.29	1	3	3	15	7.061E5	118	11.7	8.95
O82514	Adenylate kinase 4	9.90	15.85	3	2	2	8	4.629E5	246	26.9	7.36
Q9LH85	60S acidic ribosomal protein P2-3	9.47	40.87	1	2	2	5	6.116E5	115	11.7	4.68
Q8LEX2	Putative uncharacterized protein	9.21	8.93	4	1	1	3	1.850E5	168	18.3	6.18
P92549	ATP synthase subunit alpha	8.77	10.85	5	3	3	7	4.739E5	507	55.0	6.61
O80837	Remorin	8.23	21.58	1	2	2	4	2.744E5	190	21.0	8.54
Q7FY22	AT4g28750	7.81	34.55	5	1	2	9	3.686E5	110	11.7	9.79
O82204	60S ribosomal protein L28-1	7.57	25.17	1	3	3	5	4.077E5	143	15.9	10.58
B9DFQ9	AT5G37510 protein	7.31	2.55	2	1	1	3	2.452E5	745	81.1	6.64
F4III4	Mitochondrial F1F0-ATP synthase subunit Fad	6.94	7.27	2	1	1	2	1.060E6	220	25.1	9.01
P42036	40S ribosomal protein S14-3	6.88	9.33	3	2	2	3	4.229E5	150	16.2	10.59
C0Z2R2	AT4G30190 protein	6.82	3.68	12	2	2	4	1.769E5	816	89.6	8.81
O23515	60S ribosomal protein L15-1	6.57	9.31	2	1	1	2	9.010E4	204	24.2	11.44
Q9ZUX4	Uncharacterized protein At2g27730	6.37	26.55	1	2	2	2	5.671E4	113	11.9	9.64
Q8H189	40S ribosomal protein S20	5.55	22.22	3	2	2	4	1.374E5	117	13.1	9.76

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
F4HRW5	60S ribosomal protein L17-2	5.36	12.21	3	1	1	4	6.860E5	131	14.8	10.04
Q9M2Z4	Membrane steroid-binding protein 2	5.12	10.73	1	1	1	2	3.314E5	233	25.4	4.63
Q9LYN2	Ferritin-3	4.68	8.49	1	1	1	3	1.440E5	259	28.8	5.80
P51427	40S ribosomal protein S5-2	4.67	5.31	2	1	1	2	2.437E5	207	22.9	9.63
A8MQR4	60S acidic ribosomal protein P0-2	4.65	10.80	4	2	2	2	3.972E5	287	30.6	4.84
P34788	40S ribosomal protein S18	4.31	13.16	2	2	2	2	4.829E5	152	17.5	10.54
O80915	PRA1 family protein B4	4.24	6.82	1	1	1	2	8.188E4	220	23.7	7.88
Q0WLQ5	Putative tubulin alpha-2/alpha-4 chain (Fragment)	4.04	9.90	6	1	1	1	8.780E4	202	22.5	4.78
Q9M9M6	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6	3.72	27.27	1	1	1	1		110	12.2	7.43
P56771	Cytochrome f	3.34	7.81	1	1	1	1	8.026E4	320	35.3	8.29
P49688	40S ribosomal protein S2-3	3.25	5.26	2	1	1	1	7.558E4	285	30.9	10.29
F4I0N7	GA-responsive GAST1 protein-like protein	2.89	11.34	2	1	1	1	9.153E4	97	10.6	9.11
Q9S7L9	Cytochrome c oxidase subunit 6b-1	2.85	7.33	1	1	1	1	1.626E5	191	21.2	4.34
Q93VT9	60S ribosomal protein L10-1	2.83	4.55	1	1	1	1	3.545E4	220	24.9	10.49
Q93XX4	C2 domain-containing protein At1g53590	2.81	2.40	1	1	1	1	1.151E6	751	84.8	5.99
P49201	40S ribosomal protein S23-2	2.66	8.45	2	1	1	1	4.918E5	142	15.7	10.37
Q9FJW4	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4	2.51	10.39	1	1	1	1	9.815E4	154	17.1	9.64
P59230	60S ribosomal protein L10a-2	2.39	4.17	1	1	1	1		216	24.4	9.88
O64745	At2g34810	2.22	2.78	1	1	1	1	1.046E5	540	61.3	9.61
P38666	60S ribosomal protein L24-2	2.21	7.98	2	1	1	1	8.129E4	163	18.6	10.78
P56779	Cytochrome B559 subunit alpha	2.15	25.30	1	1	1	1	9.044E4	83	9.4	4.94
Q2V3P9	ATP synthase subunit d	2.15	9.84	2	1	1	1	1.737E5	122	13.8	5.71
F4J4A0	MATH domain and coiled-coil domain-containing protein At3g44790	2.11	5.25	1	1	1	1	1.898E4	324	37.2	7.94
Q9ZWT2	Cytochrome B5 isoform D	2.06	24.29	1	2	2	3	8.407E4	140	15.1	5.14
C0Z2H8	Ribosomal protein L37	2.05	12.82	4	1	1	1	3.926E4	78	8.9	10.21
H9AFL0	Ribulose bisphosphate carboxylase large chain (Fragment)	2.05	7.65	7	1	1	2		170	18.8	6.55
O22467	Histone-binding protein MSI1	2.00	4.25	1	1	1	1		424	48.2	4.82
F4HRT5	Protein CROWDED NUCLEI 1	1.81	1.50	1	1	1	1		1132	129.0	5.30
Q42564	L-ascorbate peroxidase 3, peroxisomal	0.00	6.62	1	1	1	2	1.219E5	287	31.6	6.95
O65282	20 kDa chaperonin, chloroplastic	0.00	10.67	1	1	1	1	5.794E4	253	26.8	8.88
Q42342	Cytochrome b5 isoform E	0.00	11.94	1	1	1	4	2.900E5	134	15.1	5.33
Q9FK35	Adenylate kinase 3	0.00	7.66	1	1	1	3	1.906E5	248	27.3	7.37
P42699	Plastocyanin major isoform, chloroplastic	0.00	16.77	1	1	1	1	5.244E5	167	17.0	5.20
Q9SW09	40S ribosomal protein S10-1	0.00	9.04	4	1	1	1	1.132E6	177	19.4	9.67
Q9SJ36	40S ribosomal protein S17-2	0.00	11.43	3	1	1	2	1.521E5	140	15.9	10.04
Q9M339	40S ribosomal protein S3-2	0.00	7.63	1	1	1	1	2.696E5	249	27.3	9.54
O82628	V-type proton ATPase subunit G1	0.00	30.00	1	2	2	2	6.629E4	110	12.4	6.00
Q9FMN0	Putative uncharacterized protein At5g42890	0.00	11.38	1	1	1	1		123	13.6	9.20
Q42338	AT3G48140 protein	0.00	14.77	1	1	1	1	1.141E5	88	10.0	9.45
Q9LNK3	F12K21.25	0.00	2.69	1	1	1	1		966	106.0	7.68
Q9SJV7	Cytochrome b-c1 complex subunit 6	0.00	22.58	5	1	1	1	1.412E5	62	7.4	5.77
Q8HAW5	ATPase alpha subunit (Fragment)	0.00	11.21	2	1	1	1	1.724E5	116	12.3	5.45
Q0WWY1	Regulator of chromosome condensation-like protein	0.00	4.10	2	1	1	1	2.853E6	488	51.5	5.73
O82613	T9A4.4 protein	0.00	3.92	3	1	1	1	1.348E6	485	55.3	7.59
Q56Z50	Ribosomal protein S2 (Fragment)	0.00	45.71	4	1	1	1	2.579E5	35	4.0	5.87
Q56W98	Ribosomal protein L9 (Fragment)	0.00	17.35	4	1	1	1	1.445E5	98	11.1	7.97

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9FFP4	Putative uncharacterized protein	0.00	3.15	2	1	1	1	4.588E5	571	65.6	6.55
O48728	Putative uncharacterized protein At2g26610	0.00	2.00	2	1	1	1		852	92.9	6.95

2.9 Statistically validated hits from mass spectrometry analysis of proteins obtained from the Arabidopsis stigmatic total cell protein extract.

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q41930	Polyubiquitin (Fragment)	10.49	36.84	6	1	3	5	2.232E8	38	4.3	9.70
Q56ZD8	Translation elongation factor EF-Tu, chloroplast	2.46	35.56	3	1	1	1	2.254E7	45	4.7	6.49
A0NAA9	60S ribosomal protein L39-1	2.68	27.27	2	2	2	2	9.622E7	44	5.6	12.31
Q8LPR7	AT4g33250/F17M5_10	2.68	31.15	3	2	2	2	4.329E7	61	6.8	4.94
Q56ZZ8	Putative uncharacterized protein At5g20290 (Fragment)	2.14	22.95	2	1	1	1	1.020E7	61	6.8	9.51
Q8LCW4	ABA-inducible protein-like protein	1.93	19.40	2	2	2	2	5.206E7	67	7.0	9.47
Q0WMY2	Putative uncharacterized protein At3g62530 (Fragment)	3.74	25.00	3	1	1	1	7.770E6	64	7.1	5.41
Q8LDQ8	At5g24165	5.42	17.33	1	1	1	2	9.108E6	75	7.8	10.37
Q42194	60S ribosomal protein L21	0.00	11.11	3	1	1	1	3.244E7	72	7.9	11.50
Q9FMI6	Putative uncharacterized protein	5.45	31.17	3	1	1	1	6.822E6	77	8.2	9.16
Q9FGZ9	Ubiquitin-like protein 5	2.54	15.07	1	1	1	1	2.237E7	73	8.6	8.12
Q42132	Glutathione-S-transferase (Fragment)	2.90	22.37	2	1	1	1	2.183E7	76	8.6	6.79
Q8GW48	At4g15810	15.54	16.28	1	1	1	4	4.768E8	86	9.3	4.48
Q42015	THIOREDOXINE (Fragment)	2.69	13.10	2	1	1	1	4.501E7	84	9.4	4.96
Q0WMG6	31 kDa RNA binding protein (Fragment)	3.05	16.67	3	1	1	1	1.862E7	84	9.4	5.01
Q42199	Auxin-induced mRNA (Fragment)	6.02	12.79	3	1	1	2	7.164E7	86	9.9	5.87
Q570R2	Putative SAR DNA-binding protein-1 (Fragment)	0.00	16.09	2	1	1	1	9.106E6	87	10.2	9.60
P57752	Acyl-CoA-binding domain-containing protein 6	7.59	23.91	1	2	2	4	2.840E8	92	10.4	5.24
Q8LDC9	GroES-like protein	2.10	16.49	2	1	1	1	1.081E7	97	10.5	8.50
Q7DLK3	LEA76 homolog (Fragment)	2.33	15.15	3	1	1	1	1.781E7	99	10.6	6.80
Q9XGX9	Mitochondrial import inner membrane translocase subunit TIM9	4.16	12.90	1	1	1	2	2.236E7	93	10.7	5.50
Q9SI54	Sm-like protein LSM7	1.65	7.07	1	1	1	1	4.535E7	99	10.8	5.21
Q8H7A6	Putative uncharacterized protein	3.89	12.24	3	1	1	2	7.551E7	98	10.8	7.46
P34893	10 kDa chaperonin	12.29	30.61	1	3	3	5	1.468E8	98	10.8	7.42
Q9FJJ8	Putative uncharacterized protein	1.72	6.00	2	1	1	1	6.475E6	100	10.9	10.64
A8MR50	60S ribosomal protein L34-1	6.61	16.67	3	2	2	3	5.676E7	96	11.0	11.84
Q9LXM8	60S acidic ribosomal protein P2-4	0.00	27.03	1	1	1	2	1.867E7	111	11.0	4.50
Q9SK39	Probable steroid-binding protein 3	3.51	10.00	1	2	2	3	2.146E8	100	11.0	4.88
O23095	60S acidic ribosomal protein P1-2	3.76	33.63	1	1	1	1	1.196E7	113	11.3	4.32
Q9LLR6	Non-specific lipid-transfer protein 4	6.20	11.61	1	1	1	2	1.524E9	112	11.4	8.75
C0Z2G1	AT2G36620 protein	0.00	10.00	3	1	1	1	6.183E6	100	11.4	10.74
P51407	60S acidic ribosomal protein P2-1	15.68	40.00	6	3	3	8	2.611E8	115	11.4	4.61
Q9LLR7	Non-specific lipid-transfer protein 3	6.81	13.91	1	1	1	3	8.307E8	115	11.7	8.65
Q7FY22	AT4g28750	7.21	22.73	5	1	1	4	1.744E8	110	11.7	9.79
Q9LH85	60S acidic ribosomal protein P2-3	5.91	20.87	1	2	2	2	1.486E7	115	11.7	4.68
Q42589	Non-specific lipid-transfer protein 1	7.90	13.56	1	2	2	4	3.791E8	118	11.7	8.95
Q9FNE2	Glutaredoxin-C2	23.17	59.46	2	4	4	9	5.547E7	111	11.7	7.24
Q9LDB4	Non-specific lipid-transfer protein 6	2.40	7.08	2	1	1	1	4.329E7	113	11.9	7.47
Q9S7I3	Non-specific lipid-transfer protein 2	9.75	27.97	1	2	2	3	3.230E7	118	11.9	8.97
Q8LGG0	Peptidyl-prolyl cis-trans isomerase FKBP12	10.84	11.61	1	1	1	6	1.936E8	112	12.0	6.04
Q9C6C1	RNA recognition motif-containing protein	0.00	21.62	2	1	1	1	3.432E7	111	12.0	4.55

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
P56807	30S ribosomal protein S18, chloroplastic	0.00	15.84	1	1	1	1	1.802E7	101	12.1	12.16
Q42129	O-acetylserine(Thiol) lyase (Fragment)	2.61	10.53	2	1	1	1	8.905E6	114	12.2	5.25
C0Z306	AT1G59900 protein	1.71	11.21	2	1	1	1	2.601E7	107	12.3	5.92
Q8VZ19	60S ribosomal protein L30-2	4.24	14.29	3	1	1	2	1.964E7	112	12.3	9.58
O23138	Cytochrome c-1	10.70	36.84	3	4	4	5	6.070E7	114	12.4	9.31
Q56Y83	Pyruvate decarboxylase	2.57	10.81	4	1	1	1	9.611E6	111	12.6	5.48
Q9M352	60S ribosomal protein L36-2	12.67	25.89	3	2	2	5	4.382E7	112	12.7	11.75
Q56WI6	Putative methionyl-tRNA synthetase (Fragment)	9.51	27.27	2	2	2	3	7.103E7	121	12.8	8.27
Q42058	Protein kinase inhibitor (Fragment)	1.68	7.14	5	1	1	1	7.083E7	112	12.9	5.34
Q43287	Fructose-1,6-biphosphatase (Fragment)	2.13	7.02	2	1	1	1	2.891E7	114	12.9	8.87
Q9SL42	Peptidyl-prolyl cis-trans isomerase Pin1	1.71	7.56	1	1	1	1	9.018E6	119	13.0	9.09
Q8H189	40S ribosomal protein S20	4.82	30.77	5	4	4	4	4.253E7	117	13.1	9.76
F4J3P1	60S ribosomal protein L23	6.98	17.60	4	2	2	4	5.286E7	125	13.4	10.05
Q8LGG3	Putative uncharacterized protein	0.00	13.33	2	1	1	1	9.907E6	120	13.4	10.55
O22914	Calvin cycle protein CP12-1, chloroplastic	4.29	24.19	1	1	1	1	7.515E6	124	13.5	4.93
Q9SU26	AT4g12600/T1P17_190	2.53	22.66	2	1	1	1		128	14.0	6.54
Q9M9W1	60S ribosomal protein L22-2	1.96	9.68	1	1	1	1	2.678E7	124	14.0	9.55
Q8LG89	Basic blue protein	0.00	17.05	1	1	1	1	2.452E7	129	14.0	9.50
Q9FE58	60S ribosomal protein L22-3	1.67	9.68	1	1	1	1		124	14.0	9.55
Q9LXZ1	Protein kinase C inhibitor-like protein	6.09	34.88	3	3	3	4	5.251E7	129	14.1	6.68
F4JXD5	Actin depolymerizing factor 3	5.09	11.29	2	1	1	2	1.409E7	124	14.1	5.19
Q8LC99	Adenine nucleotide alpha hydrolases-like superfamily protein	0.00	4.76	2	1	1	1	1.382E8	126	14.2	6.14
Q9C9P6	Putative ribosomal protein S9	2.62	19.70	4	2	2	2	1.211E7	132	14.3	10.08
Q94AF7	At2g07350/T13E11.12	0.00	5.47	1	1	1	1	2.363E7	128	14.6	4.93
Q9LYK9	40S ribosomal protein S26-3	2.41	30.77	3	2	2	4	3.178E7	130	14.6	11.09
A8MSB9	Glycine-rich RNA-binding protein 2	3.71	10.42	3	1	1	1	1.648E8	144	14.7	7.34
F4HRW5	60S ribosomal protein L17-2	6.13	12.21	3	1	1	2	2.260E8	131	14.8	10.04
Q8LFQ6	Glutaredoxin-C4	0.00	12.59	1	1	1	1	2.999E8	135	14.8	5.96
Q9C8W7	At1g71950	15.06	27.21	2	3	3	5	4.412E7	136	14.8	6.15
C0Z2N6	AT2G21660 protein	39.08	42.48	4	4	6	10	6.201E7	153	14.9	5.29
Q56WK2	Putative uncharacterized protein At4g14930	0.00	10.45	2	1	1	1	2.835E7	134	14.9	5.87
Q9SV91	At4g10300	2.41	9.70	2	1	1	1	2.504E8	134	14.9	8.68
O80504	Chloroplast chaperonin 10	39.77	27.34	2	4	4	10	1.245E8	139	15.0	8.48
Q42342	Cytochrome b5 isoform E	11.23	23.13	1	2	2	4	8.206E7	134	15.1	5.33
Q9C787	Putative uncharacterized protein At1g69510	1.96	8.03	1	1	1	1	1.265E7	137	15.1	5.06
P24704	Superoxide dismutase [Cu-Zn] 1	62.54	25.00	2	4	4	16	1.918E9	152	15.1	5.60
Q9M1C2	At3g60210	2.43	15.22	1	1	1	1	7.540E7	138	15.1	7.96
Q9MA77	F2J6.4 protein	7.32	17.27	2	2	2	2	3.127E7	139	15.3	4.98
Q96500	A.thaliana mRNA (orf19) from chromosome III	0.00	12.06	1	1	1	2	1.613E8	141	15.3	4.68
Q41912	50S ribosomal protein L2 (Fragment)	15.44	29.86	2	4	4	7	2.773E7	144	15.3	11.49
Q9SS17	40S ribosomal protein S24-1	4.72	15.79	2	3	3	3	2.634E7	133	15.4	10.70
P51419	60S ribosomal protein L27-3	0.00	5.19	3	1	1	1	2.780E7	135	15.6	10.26
Q0WNN2	Putative cytosolic factor protein (Fragment)	4.79	15.60	2	2	2	2	9.460E7	141	15.6	4.56
P93047	High mobility group B protein 3	3.52	9.22	2	1	1	1	3.923E7	141	15.7	5.83
Q9FNP8	40S ribosomal protein S19-3	2.14	9.79	2	1	1	1	9.019E6	143	15.7	10.21
A8MS46	Embryo defective protein 1303	2.04	8.61	2	1	1	1	4.159E6	151	15.7	9.25

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
P49201	40S ribosomal protein S23-2	2.55	21.13	2	3	3	3	5.888E7	142	15.7	10.37
O64866	Calcium-binding EF-hand-containing protein	0.00	9.86	1	1	1	1	1.018E7	142	15.8	4.51
O82204	60S ribosomal protein L28-1	0.00	11.19	1	1	1	1	9.594E6	143	15.9	10.58
Q5DWV9	Chalcone-flavonone isomerase family protein	12.88	19.73	2	3	3	4	2.656E7	147	15.9	8.25
Q9SJ36	40S ribosomal protein S17-2	0.00	8.57	4	1	1	1	1.331E7	140	15.9	10.04
Q9LXQ2	At3g44100	0.00	6.58	1	1	1	1	8.461E6	152	16.1	8.18
A8Y7S8	Z-box binding factor 3	12.90	17.61	15	3	3	5	2.208E8	142	16.1	4.34
Q96272	LEA D113 homologue type1	3.08	8.23	2	1	1	1	4.615E7	158	16.2	9.44
Q8LEH5	Putative uncharacterized protein	127.65	31.61	2	4	4	28	7.563E8	155	16.2	5.02
P42036	40S ribosomal protein S14-3	17.56	18.67	1	1	4	8	4.332E8	150	16.2	10.59
Q9SIH0	40S ribosomal protein S14-1	17.52	18.67	1	1	4	8	4.408E8	150	16.2	10.59
Q9CAX6	40S ribosomal protein S14-2	19.69	18.67	1	1	4	9	5.256E8	150	16.3	10.59
Q9MAB2	At3g05070	2.24	9.72	1	1	1	1	1.075E7	144	16.4	4.93
P49637	60S ribosomal protein L27a-3	0.00	8.22	2	1	1	1	7.873E6	146	16.4	10.59
Q9SJ44	Ubiquitin-conjugating enzyme E2 variant 1C	2.79	10.34	3	2	2	2		145	16.5	7.03
P39207	Nucleoside diphosphate kinase 1	54.95	46.98	1	8	8	18	7.446E8	149	16.5	6.79
Q03251	Glycine-rich RNA-binding protein 8	70.04	40.24	7	4	6	17	2.911E8	169	16.6	5.68
Q8LDP1	Putative uncharacterized protein	3.22	6.63	3	1	1	1	6.639E7	166	16.8	5.11
P42699	Plastocyanin major isoform, chloroplastic	20.82	16.77	1	2	2	5	7.728E7	167	17.0	5.20
Q93VR4	MLP-like protein 423	29.30	51.61	1	7	7	12	2.378E8	155	17.0	5.20
P59223	40S ribosomal protein S13-1	2.50	15.89	2	2	2	2	8.069E7	151	17.1	10.39
O82179	Glycine cleavage system H protein 2, mitochondrial	2.40	7.69	1	1	1	1	1.628E7	156	17.1	5.11
Q8RUC6	Ubiquitin-NEDD8-like protein RUB2	9.21	19.48	45	2	4	10	3.504E8	154	17.1	6.06
P61841	30S ribosomal protein S7, chloroplastic	3.40	12.26	1	1	1	2	3.148E7	155	17.3	11.28
Q9SUQ9	AT4g23680/F9D16_150	0.00	9.93	1	1	1	1	7.099E7	151	17.5	6.34
Q9SUR0	AT4G23670 protein	29.21	25.83	7	2	2	7	1.414E7	151	17.5	6.37
P34788	40S ribosomal protein S18	2.18	11.18	1	3	3	3	3.943E7	152	17.5	10.54
Q9LV66	Uncharacterized protein At5g48480	33.57	27.71	1	6	6	16	2.729E8	166	17.5	4.88
Q9FH13	Putative 4-hydroxy-4-methyl-2-oxoglutarate aldolase 3	0.00	7.23	1	1	1	1	5.072E7	166	17.6	5.55
P42733	40S ribosomal protein S11-3	2.29	15.72	2	1	2	4	1.904E8	159	17.7	10.61
Q9FFE0	Putative 4-hydroxy-4-methyl-2-oxoglutarate aldolase 2	0.00	7.23	1	1	1	1	1.676E7	166	17.8	5.58
O82355	Desiccation-related protein At2g46140	2.15	12.65	1	1	1	1	3.459E7	166	17.8	4.77
B9DI58	Nascent polypeptide-associated complex subunit beta (Fragment)	10.26	12.20	3	2	2	3	3.399E7	164	17.9	6.21
P50883	60S ribosomal protein L12-1	3.34	9.04	3	1	1	1	3.087E7	166	17.9	8.97
P16181	40S ribosomal protein S11-1	10.60	20.00	3	2	3	7	1.511E8	160	17.9	10.56
Q93ZB5	Nascent polypeptide-associated complex subunit beta	15.54	12.12	2	2	2	4	4.065E7	165	18.0	7.25
F4JUT8	Plasma-membrane associated cation-binding protein 1	9.57	18.67	2	1	1	2	1.668E8	166	18.1	6.86
Q9ZV00	Putative uncharacterized protein At2g39080	1.75	4.19	4	1	1	1	1.073E7	167	18.1	8.06
Q9SIH1	Peptidyl-prolyl cis-trans isomerase CYP18-2	2.51	14.02	1	2	2	2	4.837E6	164	18.2	8.44
Q8LEX2	Putative uncharacterized protein	25.94	17.26	4	4	4	8	9.510E8	168	18.3	6.18
Q9LE22	Probable calcium-binding protein CML27	2.08	15.88	1	2	2	2	1.092E7	170	18.3	4.35
P34790	Peptidyl-prolyl cis-trans isomerase CYP18-3	49.81	38.37	1	5	5	13	4.971E7	172	18.4	7.81
Q42406	Peptidyl-prolyl cis-trans isomerase CYP18-4	13.08	37.21	2	5	7	9	2.055E8	172	18.4	8.69
Q9SKQ0	Peptidyl-prolyl cis-trans isomerase CYP19-2	6.59	23.56	1	3	3	3	1.048E7	174	18.5	8.16
O64527	UPF0678 fatty acid-binding protein-like protein At1g79260	6.05	8.43	1	1	1	2		166	18.5	7.56
Q38900	Peptidyl-prolyl cis-trans isomerase CYP19-1	32.85	21.97	2	2	4	10	2.246E8	173	18.5	8.43

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
O23157	AT4g37300/C7A10_60	10.38	12.14	1	1	1	2	9.016E7	173	18.8	4.68
C0Z3H8	AT4G13430 protein	3.13	19.43	2	2	2	2	1.255E7	175	18.8	5.59
Q9M0Z6	Peptide methionine sulfoxide reductase B3	2.66	8.52	1	1	1	1	3.808E6	176	18.8	6.29
P31265	Translationally-controlled tumor protein homolog	0.00	7.14	3	2	2	2	2.539E7	168	18.9	4.64
Q38867	Peptidyl-prolyl cis-trans isomerase CYP19-3	15.70	28.98	2	4	5	9	1.477E8	176	18.9	7.81
Q9SZ51	Early nodulin-like protein 15	10.04	17.51	1	3	3	6	8.890E7	177	18.9	8.84
Q38896	Cold shock domain-containing protein 4	5.71	16.42	1	1	2	2	1.330E8	201	19.1	6.76
Q9C500	WPP domain-containing protein 2	0.00	7.78	1	1	1	1		180	19.1	4.53
Q41188	Cold shock protein 2	22.16	24.63	1	2	3	6	7.009E7	203	19.1	5.92
Q9FLX7	Probable NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5, mitochondrial	2.95	8.88	1	1	1	1	3.614E7	169	19.2	4.77
Q9FUA9	PRLI-interacting factor F (Fragment)	10.10	14.46	3	1	2	4	8.928E7	166	19.2	5.62
Q9XFH8	Thioredoxin F1, chloroplastic	1.99	4.49	2	1	1	1	5.627E7	178	19.3	8.92
B9DI23	AT1G47128 protein (Fragment)	6.61	10.11	6	1	2	3	4.809E7	178	19.3	6.96
Q9FJP3	50S ribosomal protein L29, chloroplastic	1.83	6.36	1	1	1	1	1.315E7	173	19.4	10.51
Q9SW09	40S ribosomal protein S10-1	0.00	9.04	4	1	1	2	1.607E8	177	19.4	9.67
Q8LGG8	Universal stress protein A-like protein	2.77	7.43	1	1	1	1	1.338E7	175	19.6	5.90
F4INP2	Uncharacterized protein	3.22	6.11	4	1	1	1	6.450E7	180	19.6	9.10
Q9FFS8	40S ribosomal protein S10-2	18.21	19.44	2	4	4	10	2.041E8	180	19.7	9.73
Q56ZW2	Adenosine kinase like protein (Fragment)	15.88	47.49	3	7	7	11	4.028E8	179	19.7	5.92
Q9ZPZ4	Putative uncharacterized protein At1g09310	10.24	11.17	1	2	2	4	9.790E7	179	19.9	5.38
Q9FIX1	AIG2-like protein	2.87	17.44	1	2	2	3	7.724E7	172	20.0	5.10
Q07488	Blue copper protein	9.91	6.63	1	1	1	9	5.522E8	196	20.0	4.82
C0Z3J9	AT5G61410 protein	1.93	5.26	2	1	1	1	1.425E8	190	20.1	6.54
P10795	Ribulose bisphosphate carboxylase small chain 1A, chloroplastic	105.32	48.89	6	5	10	37	3.512E9	180	20.2	7.71
Q9SEU8	Thioredoxin M2, chloroplastic	8.57	9.68	1	1	1	2	9.120E6	186	20.3	9.22
Q56WF9	2-oxoglutarate dehydrogenase, E1 subunit-like protein	6.02	9.55	5	1	1	2	7.092E7	178	20.3	7.75
Q56WF1	Putative uncharacterized protein At3g50370 (Fragment)	10.13	11.06	3	1	1	2	6.144E6	199	20.5	4.68
Q1G2Y5	Protease inhibitor	18.60	9.76	3	2	2	10	2.154E9	205	20.6	6.93
B9DI50	AT1G29670 protein (Fragment)	3.15	20.86	2	2	2	3	5.649E7	187	20.7	8.87
B9DGI7	AT5G51040 protein	4.31	4.89	3	1	1	2	2.802E8	184	20.7	8.75
O22875	Expressed protein	5.48	6.74	3	1	1	2	2.609E7	193	20.7	5.92
B3H5S2	Ribulose bisphosphate carboxylase small chain	61.27	48.92	2	6	11	27	3.274E9	186	20.8	8.29
Q8H0V3	Lactoylglutathione lyase	13.95	36.22	2	4	4	6	3.249E7	185	20.8	5.29
P42794	60S ribosomal protein L11-2	2.95	7.69	2	1	1	1	4.316E7	182	20.8	9.92
P42791	60S ribosomal protein L18-2	0.00	11.76	1	2	2	3	4.178E7	187	20.9	10.98
Q9C8K3	RNA-binding protein, putative; 35994-37391	0.00	7.85	3	1	1	1	9.187E7	191	21.0	4.83
Q940B0	60S ribosomal protein L18-3	0.00	18.18	1	3	3	5	8.459E7	187	21.0	10.96
Q94JX5	LIM domain-containing protein WLIM1	0.00	8.95	1	1	1	1	1.651E7	190	21.0	8.84
B0ZC56	At1g67140 (Fragment)	8.33	7.25	18	1	1	3	7.048E6	193	21.1	4.35
Q570E6	Cysteine synthase	2.85	17.17	5	1	1	2	9.369E7	198	21.1	5.44
P51418	60S ribosomal protein L18a-2	2.22	5.06	2	1	1	1	1.286E7	178	21.3	10.48
Q9SUQ8	Dirigent protein 6	6.92	11.23	1	2	2	4	7.878E7	187	21.4	8.35
Q9M7T0	Peroxiredoxin-2F, mitochondrial	4.47	5.97	1	1	1	2	9.153E6	201	21.4	8.90
Q8LAJ6	Putative uncharacterized protein	11.33	17.59	2	2	2	6	5.939E7	199	21.5	8.46
B9DHY3	AT2G37340 protein (Fragment)	10.95	14.51	3	2	2	4	3.637E7	193	21.5	10.32

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q8LDP4	Peptidyl-prolyl cis-trans isomerase CYP19-4	2.55	6.47	1	1	1	1	4.128E7	201	21.5	9.03
F4J912	60S ribosomal protein L5-1	2.30	4.74	4	1	1	1	3.199E7	190	21.5	5.86
Q9LSQ5	NAD(P)H dehydrogenase (quinone) FQR1	2.70	7.84	2	1	1	1	3.937E7	204	21.8	6.38
Q9C7Y9	At1g47970	19.74	10.10	1	2	2	5	5.194E7	198	21.8	3.57
Q9SJA6	Serine/arginine-rich splicing factor RSZ22A	2.58	6.12	1	1	1	1	2.663E7	196	21.9	11.34
Q9SP02	Peptidyl-prolyl cis-trans isomerase CYP20-1	2.01	3.92	1	1	1	1	1.825E7	204	21.9	9.06
Q9C5C8	Peptide methionine sulfoxide reductase B2, chloroplastic	19.68	17.33	1	2	2	5	9.145E7	202	22.0	8.56
O24633	Proteasome subunit beta type-2-B	3.35	9.55	2	1	1	1	1.085E7	199	22.0	6.68
Q6ICZ8	Nascent polypeptide-associated complex subunit alpha-like protein 3	6.83	7.35	1	1	1	2	7.343E6	204	22.0	4.55
Q8LD03	40S ribosomal protein S7-3	2.29	4.74	2	1	1	1	3.748E7	190	22.0	9.74
Q9LMU2	At1g17860/F2H15_8	4.02	14.29	1	1	1	1	9.483E6	196	22.1	8.78
Q8S904	Adrenodoxin-like protein 2, mitochondrial	3.14	6.09	2	1	1	2	3.665E7	197	22.1	6.86
P25864	50S ribosomal protein L9, chloroplastic	0.00	4.06	1	1	1	1	1.620E7	197	22.1	9.67
O78310	Superoxide dismutase [Cu-Zn] 2, chloroplastic	25.11	12.96	1	3	3	8	1.272E8	216	22.2	7.01
Q9ZPY1	Pyridoxine/pyridoxamine 5'-phosphate oxidase 2	7.37	23.23	1	2	2	3	1.423E7	198	22.6	7.31
B9DHT6	AT1G07750 protein (Fragment)	0.00	2.82	4	1	1	1	1.004E8	213	22.8	7.34
P51427	40S ribosomal protein S5-2	21.78	16.43	2	3	3	8	8.938E7	207	22.9	9.63
Q9LXG1	40S ribosomal protein S9-1	8.66	15.15	2	3	3	4	4.439E7	198	23.0	10.17
Q9FUS7	Glutathione S-transferase	0.00	2.90	2	1	1	1	1.989E8	207	23.0	8.21
A8MQP6	Nascent polypeptide-associated complex subunit alpha-like protein 4	3.49	7.11	2	1	1	2	4.736E7	211	23.1	4.42
Q0WSB1	Tyrosyl-tRNA synthetase-like	2.01	9.95	3	1	1	1	3.628E6	201	23.1	6.90
Q9LRX8	60S ribosomal protein L13a-2	6.83	13.11	1	1	2	4	1.252E7	206	23.4	10.35
Q9SFU1	60S ribosomal protein L13a-1	10.49	25.24	3	3	4	4	2.451E7	206	23.5	10.40
A8MRJ9	AT5G43280 protein	0.00	5.45	2	1	1	1		220	23.6	6.20
Q9FWR4	Glutathione S-transferase DHAR1, mitochondrial	32.91	36.62	2	6	6	10	3.110E8	213	23.6	5.91
Q9LU86	Peroxiredoxin Q, chloroplastic	21.69	15.74	2	4	4	7	5.973E7	216	23.7	9.51
Q944I9	At2g26740/F18A8.11	2.44	15.17	2	2	2	4	3.823E7	211	23.7	7.11
B0FUF9	Peroxidase (Fragment)	3.49	8.00	3	2	2	3	6.082E7	225	23.7	4.93
P41127	60S ribosomal protein L13-1	42.34	25.24	7	6	6	14	2.285E8	206	23.8	11.02
Q84MC2	Cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG8	5.90	11.57	1	2	2	2	1.638E8	216	23.8	5.59
O80575	6,7-dimethyl-8-ribityllumazine synthase, chloroplastic	2.60	10.57	1	1	1	2	1.765E7	227	24.0	8.35
Q9FXC0	At1g56700	2.59	5.02	1	1	1	1	6.989E6	219	24.0	6.39
Q9LX13	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase-like protein	5.21	5.48	1	1	1	2	1.702E7	219	24.1	9.26
O23443	Fumarylacetoacetate hydrolase-like protein	2.12	4.50	2	1	1	1	3.695E7	222	24.1	7.49
Q9FMB1	FrnE protein-like	5.47	5.53	1	1	1	2	9.937E7	217	24.1	6.80
Q9SCX3	Elongation factor 1-beta 2	2.84	14.29	2	2	2	2	2.360E7	224	24.2	4.56
P49693	60S ribosomal protein L19-3	1.97	4.81	4	1	1	2	4.819E8	208	24.2	11.39
O23515	60S ribosomal protein L15-1	1.77	25.98	2	5	5	5	1.090E8	204	24.2	11.44
Q9ZVL3	Nuclear transcription factor Y subunit C-3	2.17	5.99	2	1	1	1	8.131E7	217	24.3	5.12
Q8RWG8	Ran-binding protein 1 homolog b	26.19	10.14	1	1	1	6	1.001E8	217	24.4	4.88
P59230	60S ribosomal protein L10a-2	0.00	3.70	2	1	1	1	1.224E7	216	24.4	9.88
Q0WR03	Putative uncharacterized protein At2g31670	6.18	9.78	2	1	1	2	1.070E7	225	24.5	5.44
F4K9K7	Asparaginase	2.76	6.38	2	1	1	1	5.083E7	235	24.5	5.59
Q9LFF9	Soluble inorganic pyrophosphatase 4	14.12	16.67	2	2	3	6	4.978E7	216	24.6	5.47
Q42539	Protein-L-isoaspartate O-methyltransferase 1	5.17	5.22	1	1	1	2	6.820E6	230	24.6	5.77
Q41932	Oxygen-evolving enhancer protein 3-2, chloroplastic	8.47	9.13	1	1	1	3	7.878E7	230	24.6	9.72

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q940P5	Tetraspanin-19	2.09	11.31	1	1	1	1		221	24.6	7.05
Q8GYX0	MOB kinase activator-like 1B	2.43	6.05	2	1	1	2	1.281E7	215	24.7	7.52
P21216	Soluble inorganic pyrophosphatase 2	16.78	14.68	1	2	3	7	4.350E7	218	24.7	6.09
F4JD01	Proteasome subunit beta type	2.25	4.48	2	1	1	1	4.183E7	223	24.7	7.81
P92985	Ran-binding protein 1 homolog c	5.07	9.59	1	1	1	1		219	24.7	4.74
Q9FY50	50S ribosomal protein L10, chloroplastic	0.00	9.55	1	1	1	1		220	24.7	9.32
Q56ZU5	Glucose-1-phosphate adenylyltransferase	2.09	3.95	2	1	1	1	5.580E7	228	24.8	5.44
Q93VT9	60S ribosomal protein L10-1	12.04	10.00	1	2	2	6	5.129E7	220	24.9	10.49
Q39124	EEF-1beta protein	2.19	3.93	4	1	1	1	3.343E6	229	25.1	4.65
Q8LD27	Proteasome subunit beta type-6	3.80	7.30	2	1	1	2	4.005E8	233	25.1	5.55
Q8GRX2	Eukaryotic translation initiation factor 3 subunit J	2.46	4.00	1	1	1	1	5.380E6	225	25.2	4.82
P28493	Pathogenesis-related protein 5	12.90	17.57	1	3	3	4	2.744E8	239	25.2	4.98
P41916	GTP-binding nuclear protein Ran-1	5.87	11.31	4	2	2	3	4.235E8	221	25.3	6.86
F4J504	Superoxide dismutase	19.58	6.96	2	2	2	5	8.458E7	230	25.3	8.48
Q9SHG8	At1g17100	4.93	7.76	1	1	1	2	9.143E7	232	25.4	4.92
O64903	Nucleoside diphosphate kinase II, chloroplastic	0.00	12.12	3	2	2	2	1.144E8	231	25.5	8.95
F4J447	Alpha/beta-Hydrolases superfamily protein	2.71	5.08	2	1	1	1	2.367E7	236	25.5	5.20
O48646	Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial	6.34	9.91	1	3	3	3	1.816E7	232	25.6	9.35
Q94BT2	Auxin-induced in root cultures protein 12	5.02	11.11	1	2	2	2	3.136E7	252	25.6	7.90
Q9LMK7	Ran-binding protein 1 homolog a	26.91	9.65	1	1	1	5	1.471E8	228	25.6	4.94
Q9ZRW8	Glutathione S-transferase U19	0.00	7.31	1	1	1	1	4.170E6	219	25.6	6.04
B9DHH1	Nucleoside diphosphate kinase	4.53	12.61	2	2	2	2	6.920E7	238	25.6	9.09
O23708	Proteasome subunit alpha type-2-A	28.17	31.49	1	1	4	8	1.208E8	235	25.7	5.69
Q8L4A7	Proteasome subunit alpha type-2-B	24.64	31.49	1	1	4	7	1.208E8	235	25.7	5.69
O80889	At2g32520	11.09	23.01	2	3	3	5	2.758E7	239	25.9	5.47
O81149	Proteasome subunit alpha type-5-A	16.03	6.33	2	2	2	6	3.365E8	237	25.9	4.75
P52032	Phospholipid hydroperoxide glutathione peroxidase 1, chloroplastic	2.38	5.08	1	1	1	1	4.355E6	236	26.0	9.38
Q39258	V-type proton ATPase subunit E1	3.92	6.09	2	2	2	2	1.174E8	230	26.0	6.40
Q9C9C5	60S ribosomal protein L6-3	4.61	7.73	4	1	2	3	8.605E7	233	26.1	10.17
O80780	Expressed protein	5.63	5.37	1	1	1	2	2.610E7	242	26.1	5.05
Q9FZ76	60S ribosomal protein L6-1	1.68	8.15	3	1	2	2	1.700E7	233	26.1	10.10
Q8VYN9	Putative uncharacterized protein At5g54430	1.84	4.13	2	1	1	2	3.165E8	242	26.2	6.57
Q8H0X6	Cysteine proteinase inhibitor 6	0.00	8.55	1	2	2	3	2.022E8	234	26.3	6.27
C0Z300	AT2G37660 protein	4.44	7.44	2	2	2	3	2.436E7	242	26.3	5.41
O23252	Eukaryotic translation initiation factor 4E-1	15.46	7.66	1	1	1	4	1.667E8	235	26.5	5.12
B9DH86	AT5G20060 protein	3.86	7.94	2	2	2	4	6.299E7	252	26.7	6.62
Q9SYL9	50S ribosomal protein L13, chloroplastic	0.00	4.98	1	1	1	1	9.154E6	241	26.8	9.92
Q9ZW85	3-isopropylmalate dehydratase small subunit 3	4.16	9.56	2	2	2	2	4.405E7	251	26.8	6.77
O65282	20 kDa chaperonin, chloroplastic	24.18	29.25	2	5	5	8	9.616E7	253	26.8	8.88
F4KE21	Chloroplastic acetylcoenzyme A carboxylase 1	21.96	10.63	3	3	4	8	2.540E8	254	26.9	9.54
O82514	Adenylate kinase 4	7.26	15.85	3	4	4	4	1.085E8	246	26.9	7.36
Q8LDH6	Putative thaumatin	3.86	11.24	2	1	1	1	2.253E7	249	27.0	7.65
Q94EG6	Uncharacterized protein At5g02240	16.30	20.55	1	5	5	6	8.423E7	253	27.1	6.62
Q9LIL3	Aluminum induced protein with YGL and LRDR motif	2.78	5.65	1	1	1	2	8.573E7	248	27.1	6.21
P48491	Triosephosphate isomerase, cytosolic	8.80	12.60	1	2	3	4	6.321E8	254	27.2	5.50

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9LLC1	Biotin carboxyl carrier protein of acetyl-CoA carboxylase 2, chloroplastic	6.47	7.84	1	1	2	4	2.836E8	255	27.3	7.80
O81146	Proteasome subunit alpha type-6-A	14.48	13.41	3	1	3	7	2.013E8	246	27.3	5.86
Q9M339	40S ribosomal protein S3-2	5.89	12.85	3	2	3	3	1.925E7	249	27.3	9.54
O81147	Proteasome subunit alpha type-6-B	11.57	10.16	3	1	3	6	2.064E8	246	27.3	6.09
O23715	Proteasome subunit alpha type-3	4.56	12.05	1	2	2	4	5.044E8	249	27.4	6.32
Q0WP12	Thiocyanate methyltransferase 1	2.04	7.32	1	1	1	1	3.542E7	246	27.4	4.64
B9DI27	AT3G13920 protein (Fragment)	0.00	8.44	9	1	1	1	3.539E7	237	27.5	6.05
O81148	Proteasome subunit alpha type-4-A	6.89	7.60	2	2	2	3	2.616E8	250	27.5	7.11
Q9FG81	Aluminium induced protein with YGL and LRDR motifs	7.94	23.11	2	4	4	8	8.839E7	251	27.5	6.87
Q9SIP7	40S ribosomal protein S3-1	11.52	12.00	3	2	3	4	2.106E7	250	27.5	9.54
F4HU93	L-ascorbate peroxidase 1	22.21	24.50	3	5	6	13	2.155E8	249	27.5	6.29
Q9SZG5	Possible apospory-associated like protein(Fragment)	2.18	5.44	4	1	1	1	5.254E6	239	27.6	9.55
Q7DLR9	Proteasome subunit beta type-4	39.00	28.46	1	6	6	10	4.840E7	246	27.6	6.55
F4K5C7	40S ribosomal protein S4	2.94	14.75	6	2	2	2	3.483E7	244	27.7	9.99
F4KGV2	14-3-3-like protein GF14 lambda	2.70	12.60	7	2	2	2	1.213E7	246	27.7	4.89
O80840	Phosphomannomutase	5.31	4.47	1	1	1	2	1.894E7	246	27.7	5.54
O65386	F12F1.20 protein	2.93	5.10	2	1	1	1	1.320E7	255	27.7	5.20
Q93ZC5	Allene oxide cyclase 4, chloroplastic	7.30	7.48	1	1	1	2	3.686E7	254	27.8	9.07
P46286	60S ribosomal protein L8-1	38.91	35.27	3	3	8	17	1.483E8	258	27.8	10.90
O81835	AT4G27320 protein	1.61	5.02	6	1	1	1	1.957E8	259	27.9	6.43
Q42064	60S ribosomal protein L8-3	23.72	23.26	2	1	6	10	1.419E8	258	27.9	10.83
Q9SE96	GEM-like protein 1	0.00	10.04	1	1	1	2	3.610E7	259	27.9	6.64
B9DG07	AT3G45300 protein (Fragment)	0.00	4.69	2	1	1	1	1.528E7	256	28.0	7.71
F4IX28	Peptidyl-prolyl cis-trans isomerase	5.42	4.63	3	1	1	2	1.485E7	259	28.1	8.75
Q42029	Oxygen-evolving enhancer protein 2-1, chloroplastic	21.50	9.51	1	3	3	8	1.956E8	263	28.1	7.39
P51430	40S ribosomal protein S6-2	18.86	25.30	2	2	5	8	1.477E8	249	28.1	10.83
P60040	60S ribosomal protein L7-2	10.02	16.12	6	3	3	4	3.342E7	242	28.2	9.94
Q39101	Ferritin-1, chloroplastic	9.10	6.27	1	1	1	3	2.140E8	255	28.2	6.11
Q56X90	Carbonic anhydrase	29.83	14.67	3	1	4	8	5.450E7	259	28.2	5.40
Q9STG3	Putative cullin-like protein 4	2.59	8.50	1	1	1	1		247	28.3	7.65
Q9LHT0	Tropinone reductase homolog At5g06060	0.00	7.95	1	1	1	1	2.786E7	264	28.3	7.69
F4II65	TGF-beta receptor interacting protein 1	3.61	6.69	3	1	1	1	6.762E6	254	28.3	7.50
Q8L9J9	Probable carbohydrate esterase At4g34215	16.93	8.85	1	2	2	5	2.731E7	260	28.3	5.97
O48549	40S ribosomal protein S6-1	26.65	25.20	3	2	5	8	1.710E8	250	28.3	10.61
Q56XG1	Putative uncharacterized protein At5g12410	3.79	4.28	2	1	1	2	7.793E6	257	28.5	5.03
Q9LK22	Plant UBX domain-containing protein 1	0.00	3.59	1	1	1	1	8.492E6	251	28.5	6.73
Q8LE52	Glutathione S-transferase DHAR3, chloroplastic	2.23	14.73	2	3	3	3	3.627E7	258	28.5	7.74
Q8LC65	Expansin-like 1 (At-EXPL1) (Ath-ExpBeta-2.1)	2.90	5.66	2	1	1	1	8.276E7	265	28.7	7.99
O24496	Hydroxyacylglutathione hydrolase cytoplasmic	4.40	15.12	1	3	3	3	5.013E7	258	28.8	6.38
Q9ZUC2	Beta carbonic anhydrase 3	1.70	2.71	1	1	1	1	1.323E7	258	28.8	6.98
F4I5Y8	Eukaryotic translation initiation factor 3 subunit H	23.52	7.60	2	2	2	6	1.020E8	250	28.9	5.15
Q8L428	EF-Hand containing protein-like	2.67	12.45	2	2	2	5	4.531E7	265	29.0	4.77
Q9M9S3	Single-stranded DNA-binding protein WHY1, chloroplastic	8.86	12.55	1	2	2	3	1.070E7	263	29.0	9.38
O49678	Putative uncharacterized protein T18B16.170	2.28	5.05	2	1	1	2	8.840E6	277	29.1	7.91
Q96291	2-Cys peroxiredoxin BAS1, chloroplastic	39.08	15.41	1	1	2	10	3.039E8	266	29.1	7.44

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
P49692	60S ribosomal protein L7a-1	1.96	13.23	3	3	3	6	1.239E7	257	29.1	10.13
O49499	Caffeoyl-CoA O-methyltransferase 1	10.67	16.22	4	3	3	4	3.920E7	259	29.1	5.29
Q01525	14-3-3-like protein GF14 omega	0.00	9.65	12	1	2	2	1.994E7	259	29.1	4.79
B9DHK8	AT1G61580 protein (Fragment)	2.69	4.67	6	1	1	1	3.577E7	257	29.2	9.89
O65484	Putative uncharacterized protein AT4g23330	0.00	1.93	9	1	1	1	5.595E7	259	29.2	5.24
Q9LFT6	Alpha-hydroxynitrile lyase	4.79	4.26	1	1	1	2	3.092E7	258	29.2	6.07
Q9ZW35	Proliferating cell nuclear antigen 2	1.63	3.03	1	1	1	1	1.380E7	264	29.2	4.72
Q96266	Glutathione S-transferase F8, chloroplastic	1.85	9.13	1	2	2	4	2.769E7	263	29.2	8.50
Q9S726	Probable ribose-5-phosphate isomerase 3, chloroplastic	2.05	6.52	1	1	1	2	9.010E7	276	29.3	6.02
B9DFC0	AT4G08870 protein	31.02	23.19	3	1	4	12	2.123E8	263	29.3	7.12
Q8LC68	NAP1-related protein 2	1.73	3.53	2	1	1	1	3.035E7	255	29.3	4.30
Q9M3C2	Oxidoreductase-like protein (Fragment)	8.26	9.89	15	4	4	4	2.947E7	273	29.4	8.22
F4HST2	Dehydrin ERD10	7.79	9.27	2	2	2	3	2.469E8	259	29.4	5.47
Q8LEA5	2-cys peroxiredoxin-like protein	23.53	15.13	2	1	2	6	1.036E8	271	29.5	6.00
Q94C69	Cold shock domain-containing protein 3	15.24	19.93	1	3	3	5	3.979E7	301	29.5	7.42
Q8L768	AT1G78150 protein	0.00	6.93	3	2	2	2	3.395E7	274	29.7	6.58
P25873	50S ribosomal protein L15, chloroplastic	3.00	3.97	1	1	1	1	1.500E7	277	29.7	10.77
Q66GR6	Single-stranded DNA-binding protein WHY3, chloroplastic	4.53	5.97	1	1	1	1	1.177E7	268	29.7	9.52
Q96300	14-3-3-like protein GF14 nu	4.51	10.57	12	1	2	2	1.505E7	265	29.8	4.82
Q9CAV0	40S ribosomal protein S3a-1	3.70	5.34	2	2	2	2	3.528E7	262	29.8	9.76
P31168	Dehydrin COR47	21.59	15.85	4	3	3	8	4.390E8	265	29.9	4.77
P42643	14-3-3-like protein GF14 chi	2.44	8.24	14	2	3	4	2.175E7	267	29.9	4.81
Q7XJ55	At5g59490	0.00	4.89	2	1	1	1	2.637E7	266	30.0	4.78
Q9LK01	At3g24420	1.97	2.93	1	1	1	1	3.648E7	273	30.0	5.49
O65639	Cold shock protein 1	25.55	30.77	1	6	6	8	4.668E7	299	30.1	7.84
Q2V4Q4	50S ribosomal protein L4	2.87	5.76	3	1	1	2	4.938E7	278	30.1	9.26
Q9M6K2	Isopentenyl pyrophosphate:dimethyllallyl pyrophosphate isomerase (Fragment)	18.16	16.09	4	4	4	9	3.535E7	261	30.2	5.71
P42645	14-3-3-like protein GF14 upsilon	25.27	8.58	12	1	2	9	2.166E8	268	30.2	4.81
O49195	Vegetative storage protein 1	44.32	42.59	4	10	10	17	2.345E8	270	30.2	5.67
Q9LNN2	Lectin-like protein At1g53070	3.20	6.25	1	2	2	3	3.521E7	272	30.4	8.50
Q9M1X0	Ribosome-recycling factor, chloroplastic	10.50	6.91	1	1	1	2	3.093E7	275	30.4	9.44
P43286	Aquaporin PIP2-1	1.98	4.53	1	1	1	1	1.121E7	287	30.5	8.40
P34066	Proteasome subunit alpha type-1-A	18.17	15.11	2	3	3	5	4.464E8	278	30.5	5.08
O65220	Peptidyl-prolyl cis-trans isomerase CYP28, chloroplastic	4.90	10.32	1	2	2	2	6.670E6	281	30.5	7.27
A3FBB7	Disease resistance protein Rpp8-like protein (Fragment)	0.00	4.91	4	1	1	1	6.446E7	265	30.6	7.36
A8MQR4	60S acidic ribosomal protein P0-2	0.00	5.23	3	1	1	2	3.701E7	287	30.6	4.84
Q41969	Eukaryotic translation initiation factor 2 subunit beta	13.03	5.22	1	2	2	7	3.033E7	268	30.6	7.20
F4JD59	UV excision repair protein RAD23C	10.96	5.35	3	2	2	3	3.690E7	299	30.6	5.02
C0Z2E9	AT5G35630 protein	4.55	5.04	2	1	1	2	3.686E7	278	30.7	7.71
Q9LSV0	Glyoxylate/succinic semialdehyde reductase 1	2.35	7.61	1	2	2	2	1.674E7	289	30.7	6.13
Q9ZUU4	RNA-binding protein CP29B, chloroplastic	7.77	7.96	1	1	1	2	2.698E8	289	30.7	5.16
Q9C8L2	Fatty-acid-binding protein 3	1.75	2.44	1	1	1	1	6.456E6	287	30.7	8.41
G1JSH2	At1g58380	6.67	15.85	5	4	4	5	4.901E7	284	30.8	10.26
Q94CE4	Beta carbonic anhydrase 4	7.23	4.64	1	2	2	3	2.508E7	280	30.8	7.09
Q94B60	ATP-dependent Clp protease proteolytic subunit 4, chloroplastic	1.63	2.40	1	1	1	1	4.600E7	292	31.5	5.55

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9LHH7	Bifunctional protein FolD 2	28.83	24.08	1	4	4	10	6.520E7	299	31.6	8.16
O80507	Putative casein kinase II subunit beta-4	2.14	3.18	1	1	1	1	9.827E6	283	31.6	5.54
Q9AV97	2-dehydro-3-deoxyphosphooctonate aldolase 1	3.93	14.14	4	3	3	3	7.893E6	290	31.6	6.79
F4J9Y8	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	2.46	5.76	2	1	1	1	2.737E7	278	31.7	5.64
Q9FGS0	RNA-binding protein CP31B, chloroplastic	8.75	4.84	1	1	1	3	5.249E7	289	31.8	4.96
B9DGT0	Lactoylglutathione lyase	12.89	14.49	5	4	4	6	8.128E8	283	32.0	5.27
B9DG17	40S ribosomal protein SA	11.67	11.90	2	2	2	5	7.906E7	294	32.0	5.15
Q8LDF2	2-nitropropane dioxygenase-like protein	2.23	5.46	2	1	1	1	2.032E7	293	32.1	5.85
C0Z3F3	AT4G13940 protein	36.29	18.21	10	6	6	25	4.639E8	291	32.1	7.94
F4IT21	Uncharacterized protein	1.64	2.77	1	1	1	1	3.255E7	289	32.1	6.71
Q9M336	Uracil phosphoribosyltransferase, chloroplastic	2.17	4.05	1	1	1	1	9.217E6	296	32.2	8.95
Q8S8F8	GLABRA2 expression modulator	31.85	11.37	1	3	3	10	1.494E8	299	32.2	5.41
Q8LDZ6	Dual transcription unit and alternative splicing protein GLAUCE	17.34	22.38	2	7	7	12	5.381E8	286	32.2	6.01
B9DFU4	AT1G20020 protein (Fragment)	25.91	23.81	3	5	5	9	6.848E7	294	32.3	8.76
F4HRK0	Hydroxyacylglutathione hydrolase 3	16.54	8.16	2	1	1	4	1.010E8	294	32.3	7.01
F4KC24	Xylose isomerase	16.77	5.92	4	1	1	4	2.369E7	287	32.4	7.75
Q42538	Serine acetyltransferase 5	8.59	11.86	1	2	2	3	3.639E7	312	32.7	7.18
Q9M2E2	(+)-neomenthol dehydrogenase	2.24	5.07	2	1	1	1	2.711E7	296	32.8	5.47
Q9FEF8	Probable mediator of RNA polymerase II transcription subunit 36b	2.72	6.17	2	1	1	1	9.713E6	308	32.8	10.14
Q9ZUH5	Short-chain dehydrogenase/reductase 2b	0.00	7.09	1	1	1	1	1.308E7	296	32.8	5.77
Q8RWU7	Plant UBX domain-containing protein 4	13.68	5.28	1	1	1	4	1.128E7	303	32.9	5.95
F4IKM1	Tropinone reductase homolog At2g29340	7.00	5.86	1	1	1	3	1.259E7	307	32.9	5.67
A0A097PL68	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase family protein (Fragment)	2.83	4.64	2	1	1	1	1.430E8	302	33.0	6.64
Q8LDE9	GrpE protein homolog	1.66	3.31	2	1	1	1	7.631E6	302	33.1	6.73
F4KFJ2	Target of Myb protein 1	2.65	4.38	2	1	1	1	2.839E7	297	33.3	4.89
Q9SKP6	Triosephosphate isomerase, chloroplastic	15.12	13.33	2	3	4	5	8.751E7	315	33.3	7.83
C0Z2K9	AT1G16080 protein	5.21	3.88	2	2	2	2	7.109E7	309	33.3	6.76
Q9LXC9	Soluble inorganic pyrophosphatase 6, chloroplastic	2.34	4.00	1	1	1	2	5.820E7	300	33.4	6.01
Q9LPL2	F24J8.7 protein	0.00	2.96	2	1	1	1	1.009E7	304	33.4	7.53
Q94CE3	Beta carbonic anhydrase 5, chloroplastic	2.41	5.65	1	1	1	1	3.696E7	301	33.4	8.41
Q9FGE4	Putative uncharacterized protein At5g24460	7.13	5.67	1	1	1	2	7.720E6	300	33.4	9.17
P33207	3-oxoacyl-[acyl-carrier-protein] reductase, chloroplastic	33.64	21.00	1	6	6	11	3.110E8	319	33.5	9.41
Q9LQ04	Bifunctional/ dTDP-4-dehydrorhamnose reductase	2.26	3.32	5	1	1	1	3.386E7	301	33.6	5.97
A2RVS6	Serine/arginine-rich splicing factor SR34A	2.69	4.00	1	1	1	1	2.593E7	300	33.6	10.83
O49629	Probable plastid-lipid-associated protein 2, chloroplastic	3.68	2.26	1	1	1	2	5.464E7	310	33.6	5.83
B9DHQ7	AT4G01850 protein (Fragment)	44.27	20.92	2	2	7	22	9.387E8	306	33.6	8.02
P52577	Isoflavone reductase homolog P3	4.62	12.90	1	4	4	5	3.908E7	310	33.7	5.94
Q0WU71	Putative uncharacterized protein At1g44835	6.28	6.19	3	1	1	3	2.229E7	307	33.7	8.43
O82299	Putative chloroplast RNA binding protein	0.00	2.92	1	1	1	1	3.963E7	308	33.8	5.64
P47998	Cysteine synthase 1	17.49	15.84	4	3	3	5	3.543E7	322	33.8	6.14
Q9LZ82	Protein BTR1	7.89	10.86	1	3	3	4	1.572E7	313	33.8	6.01
A8MRW5	Oxidoreductase, zinc-binding dehydrogenase family protein	13.90	25.08	3	4	4	6	2.690E7	319	33.8	7.28
Q9ZNR6	Pyridoxal 5'-phosphate synthase-like subunit PDX1.2	11.85	5.41	1	1	1	3	2.896E7	314	33.8	5.72
Q9SXJ6	ATP-dependent Clp protease proteolytic subunit 3, chloroplastic	5.45	9.71	1	1	1	2		309	33.9	7.72

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q8W1X2	Pyridoxal kinase	0.00	3.88	2	1	1	1	2.640E7	309	34.0	5.83
D3K046	Auxin response factor 5 (Fragment)	0.00	6.07	13	1	1	1	1.867E5	313	34.1	4.61
Q9FLC0	Peroxidase 52	13.92	11.73	3	4	4	5	2.546E7	324	34.2	8.31
Q9SZQ5	WD repeat-containing protein VIP3	1.62	3.12	1	1	1	1	2.076E7	321	34.2	6.29
F4K5T2	Bifunctional cystathionine gamma-lyase/cysteine synthase	0.00	3.72	3	1	1	1	1.940E7	323	34.3	5.78
F4HNZ6	Glyceraldehyde-3-phosphate dehydrogenase (NADP+) (Phosphorylating)	8.74	12.93	4	1	3	4	8.281E7	317	34.3	6.64
Q9ASQ2	AT5g45420/MFC19_9	9.80	6.15	2	1	1	2	7.150E6	309	34.3	8.40
F4J9Z1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	4.91	5.39	4	1	1	2	1.519E7	297	34.4	5.08
Q93ZH5	At1g21080/T22I11_9	6.84	5.59	2	1	1	2	8.861E6	304	34.4	6.81
Q9C8Q2	Pectinesterase, putative, 5' partial; 91413-90223 (Fragment)	5.22	4.55	4	1	1	2	8.566E6	308	34.4	9.51
A8MQY4	Carbonic anhydrase	17.97	12.26	4	1	4	6	3.702E7	310	34.4	7.90
Q0WM74	Methionine S-methyltransferase (Fragment)	11.83	15.63	2	2	2	5	1.623E7	320	34.4	7.58
Q9LV09	Protein BOBBER 1	12.34	6.91	1	1	1	3	4.678E7	304	34.5	5.34
Q9LT42	Genomic DNA, chromosome 3, P1 clone: MOE17	1.75	3.16	2	1	1	1	1.344E7	316	34.6	5.76
Q9FKG8	Putative quinone oxidoreductase	5.49	3.40	2	1	1	3	1.519E8	324	34.7	6.93
O80944	Aldo-keto reductase family 4 member C8	0.00	3.86	1	1	1	1	3.727E7	311	34.7	6.99
Q9C6U3	Putative uncharacterized protein T8G24.2 (Fragment)	0.00	2.79	2	1	1	1	5.059E7	323	34.7	7.59
A8MR76	Putative DNA repair protein RAD23-4	25.89	10.84	3	3	3	11	8.170E7	332	34.7	4.92
Q9ZPQ3	D-aminoacyl-tRNA deacylase	0.00	4.10	1	1	1	2	1.184E7	317	34.7	6.54
Q944A5	AT3g01590/F4P13_13	12.80	10.13	5	2	2	4	1.496E7	306	34.7	6.95
Q8GXH6	3-isopropylmalate dehydrogenase	8.63	8.05	6	1	2	3	1.679E7	323	34.8	5.20
Q9LW52	Genomic DNA, chromosome 3, P1 clone: MLM24	36.83	51.77	1	7	7	12	2.280E8	452	34.8	10.81
F4KGH1	Annexin	0.00	8.28	2	2	2	2	4.738E7	302	34.9	6.14
Q42578	Peroxidase 53	2.29	2.09	2	1	1	1	5.287E7	335	35.0	4.96
Q9M338	Aldo-keto reductase family 4 member C11	0.00	6.03	1	1	1	1	4.044E6	315	35.0	6.95
Q9SN79	Alpha/beta-hydrolase domain-containing protein	8.03	4.21	1	1	1	3	8.624E7	309	35.0	7.90
B9DH39	AT2G25670 protein	11.02	10.06	4	1	1	2	1.711E7	318	35.1	5.12
Q0PGJ6	Aldo-keto reductase family 4 member C9	7.08	6.67	1	2	2	5	1.414E7	315	35.1	8.12
P23321	Oxygen-evolving enhancer protein 1-1, chloroplastic	38.45	20.48	2	4	4	10	2.552E8	332	35.1	5.66
W8PV22	Glycosyltransferase (Fragment)	8.19	4.81	2	1	1	3	2.281E7	312	35.2	6.46
Q94BR8	Aldose 1-epimerase	7.66	13.62	2	2	2	3	1.005E7	323	35.2	6.60
A8MS79	Aldolase-type TIM barrel family protein	0.00	5.06	3	2	2	3	8.978E6	316	35.2	5.82
Q9SID0	Probable fructokinase-1	5.34	8.62	3	3	3	3	6.923E7	325	35.3	5.49
Q9SUU6	Putative uncharacterized protein AT4g32460	3.03	4.00	3	1	1	1	2.687E7	325	35.3	8.75
Q8LB95	Putative ubiquitin fusion-degradation protein	2.03	3.45	3	1	1	1	5.872E7	319	35.3	6.19
Q9SZE1	Probable 3-hydroxyisobutyrate dehydrogenase-like 1, mitochondrial	12.13	10.48	1	3	3	4	3.857E7	334	35.3	8.40
P59120	Peroxidase 58	4.62	10.03	2	2	2	2	2.975E7	329	35.4	5.31
Q9LVC5	Apospory-associated protein C	9.05	4.81	1	2	2	3	1.367E7	312	35.4	6.04
Q9M9S0	Zinc-finger homeodomain protein 4	7.18	6.73	1	1	1	4	1.658E7	312	35.5	7.59
Q56WR2	Putative uncharacterized protein	3.51	4.43	2	1	1	1	6.971E6	316	35.5	5.01
Q9FYE1	Metacaspase-9	2.51	4.00	1	1	1	1	1.049E7	325	35.5	6.24
Q9SMN0	Probable carboxylesterase 12	21.23	6.17	1	1	1	8		324	35.5	5.43
F4JU04	Glutathione S-transferase family protein	6.43	10.54	4	3	3	3	2.295E7	313	35.5	5.97
P93819	Malate dehydrogenase, cytoplasmic 1	93.62	37.05	4	7	11	31	6.926E8	332	35.5	6.55

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9SUT2	Peroxidase 39	28.73	19.94	1	5	5	10	1.387E8	326	35.6	6.98
F4IL52	Protein disulfide-isomerase like 2-1	21.95	17.96	6	5	5	9	2.028E8	323	35.6	6.00
Q945P1	At2g39050/T7F6.22	4.99	4.42	1	1	1	3	1.479E8	317	35.6	6.70
B3H6F9	Calcium-independent ABA-activated protein kinase	2.73	4.14	9	1	1	1	6.029E6	314	35.6	4.97
P57106	Malate dehydrogenase, cytoplasmic 2	44.67	25.30	3	3	7	14	2.443E8	332	35.7	6.79
Q8LPN7	E3 ubiquitin-protein ligase RING1-like	9.34	9.76	1	2	2	8	6.166E6	328	35.7	4.51
Q9M7E7	Leucine-rich repeat protein FLR1	0.00	6.79	3	1	1	1		324	35.7	8.43
Q39061	RNA-binding protein CP33, chloroplastic	12.85	3.34	1	2	2	8	1.666E8	329	35.7	4.64
Q9ZP06	Malate dehydrogenase 1, mitochondrial	25.84	18.18	3	3	3	12	4.130E8	341	35.8	8.35
Q8LCW6	Similar to late embryogenesis abundant proteins	0.00	6.48	2	1	1	1	9.763E7	324	35.9	4.87
P83291	NADH-cytochrome b5 reductase-like protein	5.08	5.18	1	2	2	4	3.721E7	328	36.0	8.69
Q9SJM7	Uridine nucleosidase 1	2.15	3.87	1	1	1	1		336	36.1	5.19
P68209	Succinyl-CoA ligase [ADP-forming] subunit alpha-1, mitochondrial	22.19	19.60	3	5	5	12	1.272E8	347	36.1	8.27
Q9LQ87	T1N6.10 protein	0.00	2.50	4	1	1	1	1.701E7	320	36.1	9.74
Q9SYT0	Annexin D1	12.22	13.56	2	3	3	4	1.121E8	317	36.2	5.38
O64640	Probable carboxylesterase 8	7.39	9.42	1	2	2	2	6.221E7	329	36.4	6.46
F4I7M5	Spermidine synthase 1	12.60	4.28	3	2	2	4	8.365E7	327	36.5	5.25
O23016	Probable voltage-gated potassium channel subunit beta	9.59	9.76	1	1	1	3	6.164E7	328	36.5	7.42
Q541D6	Polygalacturonase inhibiting protein	0.00	4.60	2	1	1	1	1.226E8	326	36.6	8.81
Q9FVC4	Aspartate-semialdehyde dehydrogenase (Fragment)	0.00	6.18	3	1	1	4	1.299E8	340	36.6	5.58
Q38814	Thiamine thiazole synthase, chloroplastic	15.13	22.64	2	7	7	9	5.113E7	349	36.6	6.23
Q9SJZ2	Peroxidase 17	0.00	3.95	1	1	1	1	6.962E7	329	36.6	5.22
Q0WLN1	Putative uncharacterized protein At1g15410	0.00	6.36	2	1	1	2		330	36.7	7.14
B1GV75	Flavone synthase (Fragment)	2.88	4.94	5	1	1	1	8.659E7	324	36.8	5.76
Q29Q26	Ankyrin repeat-containing 2B	2.39	2.62	1	1	1	1	5.015E7	344	36.9	4.48
P25858	Glyceraldehyde-3-phosphate dehydrogenase GAPC1, cytosolic	65.06	32.84	7	10	10	24	6.588E8	338	36.9	7.12
F4IBT7	Gamma-glutamyl hydrolase 2	4.01	9.01	2	1	1	1	4.004E7	333	36.9	7.14
Q9SAR5	Ankyrin repeat domain-containing protein 2	0.00	2.92	2	1	1	1	3.217E6	342	37.0	4.58
Q9FLH8	Probable fructokinase-7	4.20	2.92	2	1	1	2	1.846E7	343	37.0	5.15
Q8H1Q2	Cytosolic Fe-S cluster assembly factor NBP35	1.72	2.57	1	1	1	1	1.147E7	350	37.3	4.89
P46637	Arginase 1, mitochondrial	25.81	17.25	2	1	4	9	8.030E7	342	37.3	6.55
F4IF83	Putative DNA repair protein RAD23-1	3.81	6.55	3	1	1	1	4.836E6	351	37.3	4.48
Q24JL3	Thiosulfate/3-mercaptopyruvate sulfurtransferase 2	2.84	6.14	1	1	1	2	7.578E6	342	37.4	6.13
F4JFY4	L-ascorbate peroxidase S	23.30	18.79	3	4	6	11	1.662E8	346	37.4	9.06
O80574	4-hydroxy-tetrahydrodipicolinate reductase 1, chloroplastic	7.53	6.63	1	1	1	2	1.861E7	347	37.5	6.46
Q42546	SAL1 phosphatase	1.79	8.78	1	2	2	2	1.366E7	353	37.5	5.17
B3H533	AAA-type ATPase family protein	0.00	5.67	3	1	1	1	2.151E7	335	37.5	8.21
Q8GXQ8	Putative inosine-5'-monophosphate dehydrogenase	23.91	7.43	2	1	1	7	1.015E8	350	37.7	7.84
Q96512	Peroxidase 9	14.42	10.98	1	2	2	3	3.276E7	346	37.7	7.31
Q93ZN2	Probable aldo-keto reductase 4	52.67	29.57	5	7	7	23	2.597E8	345	37.9	6.30
Q8LDQ7	Nuclear RNA binding protein A-like protein	80.08	22.97	5	6	6	23	1.278E8	357	37.9	8.53
C0Z3C3	AT1G75330 protein	2.32	9.25	2	3	3	3	5.485E7	346	38.0	6.90
Q93YW7	Cardiolipin synthase (CMP-forming), mitochondrial	1.99	4.99	1	1	1	1	1.077E7	341	38.0	10.14
Q94AM2	Putative dihydrolipoamide S-acetyltransferase (Fragment)	41.81	22.49	2	6	6	17	3.151E8	369	38.1	7.47
P32961	Nitrilase 1	19.83	14.16	7	4	4	7	1.117E8	346	38.1	6.28
Q6NPM8	Bifunctional phosphatase IMPL2, chloroplastic	3.61	4.05	1	1	1	1	1.075E7	346	38.2	6.43

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
P42738	Chorismate mutase 1, chloroplastic	2.05	4.12	1	1	1	2	5.921E6	340	38.2	5.92
Q0WNH3	Putative uncharacterized protein At1g54890	3.57	7.20	2	1	1	1	1.670E8	347	38.3	7.94
Q9SJQ9	Fructose-bisphosphate aldolase	57.98	30.45	7	2	8	20	5.004E8	358	38.4	7.39
Q9FZA2	Non-classical arabinogalactan protein 31	2.62	5.29	1	1	1	2	1.290E7	359	38.5	10.17
Q9LF98	Fructose-bisphosphate aldolase	82.63	53.91	7	7	13	30	7.127E8	358	38.5	6.46
Q9FIE8	UDP-glucuronic acid decarboxylase 3	17.20	18.42	3	3	3	6	2.830E7	342	38.5	8.12
Q9FFD2	Probable UDP-arabinopyranose mutase 5	4.44	5.75	1	2	2	2	1.089E8	348	38.6	5.21
Q9LVI8	Glutamine synthetase cytosolic isozyme 1-3	11.27	12.15	6	3	3	6	2.282E8	354	38.6	6.06
C0Z3E9	AT4G38220 protein	4.76	5.46	5	3	3	3	3.198E7	348	38.7	6.89
Q8LEV8	Putative eukaryotic translation initiation factor 2 alpha subunit, eIF2	2.45	8.72	3	2	2	2	1.636E7	344	38.7	5.16
Q3E9G3	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.00	4.69	2	1	1	1		341	38.8	8.05
Q9SMU8	Peroxidase 34	30.59	13.88	8	5	5	10	3.369E8	353	38.8	7.64
O23593	AT4g17520/dl4795w	6.77	9.72	2	3	3	3	2.215E7	360	38.9	8.75
P24101	Peroxidase 33	0.00	1.98	1	1	1	1	1.587E7	354	38.9	6.87
Q9FK51	ADP-glucose phosphorylase	16.31	6.84	1	2	2	5	1.733E7	351	39.0	6.68
F4JWF3	DEAD-box ATP-dependent RNA helicase 15	2.15	3.78	4	1	1	2	1.484E7	344	39.1	6.84
P48523	Cinnamyl alcohol dehydrogenase 4	1.90	1.92	1	1	1	2	3.217E7	365	39.1	5.55
Q94BN2	Spermine synthase	17.02	4.46	2	1	1	6	2.648E8	359	39.2	5.92
Q9FJ95	Sorbitol dehydrogenase	0.00	4.12	1	1	1	1	9.550E6	364	39.2	5.97
Q8H7D1	Putative uncharacterized protein	4.66	5.49	2	1	1	2	1.136E7	346	39.2	5.92
Q94A80	AT5g41970/MJC20_7	2.70	5.49	3	2	2	3	4.608E7	346	39.3	5.92
F4JGF4	Ferredoxin--NADP reductase, root isozyme 1	9.03	8.00	2	3	3	5	3.546E7	350	39.3	8.22
Q94K85	Cathepsin B-like cysteine protease	11.81	5.57	2	1	1	3	2.421E7	359	39.4	6.18
Q8LBA4	Zinc-binding dehydrogenase, putative	0.00	3.55	2	1	1	3	1.948E7	366	39.4	9.03
A8MS37	Peroxisomal (S)-2-hydroxy-acid oxidase GLO1	2.09	9.17	9	2	2	3	1.149E8	360	39.4	9.50
Q9FGE2	Beta-galactosidase related protein	0.00	3.71	1	1	1	1	5.134E7	350	39.4	8.97
F4J5J9	Cysteine proteinases superfamily protein	10.49	3.36	5	1	1	4	2.550E7	357	39.5	6.68
O82359	Sphingoid long-chain bases kinase 2, mitochondrial	0.00	4.67	1	1	1	1	4.585E7	364	39.6	8.48
Q9LM66	Xylem cysteine proteinase 2	9.77	10.96	1	3	3	3	5.721E7	356	39.7	5.29
Q9SUT5	Protein SGT1 homolog B	2.78	4.19	1	1	1	1	2.147E7	358	39.7	5.11
Q9FI36	At5g48020	14.28	6.20	1	1	1	4	1.453E8	355	39.8	5.59
Q9LT39	Leucine-rich repeat-containing protein	0.00	1.64	1	1	1	1	7.820E6	365	39.8	8.41
Q0WMQ0	Putative uncharacterized protein At4g05150 (Fragment)	2.26	3.60	4	1	1	1	3.970E6	361	39.9	5.48
F4JEX6	Uncharacterized protein	0.00	2.19	2	1	1	1	4.873E7	366	40.2	8.12
Q9FKW6	Ferredoxin--NADP reductase, leaf isozyme 1, chloroplastic	10.19	6.39	3	3	3	4	5.486E7	360	40.3	8.13
O22940	At2g41800/T11A7.10	2.66	5.41	1	1	1	1	1.075E7	370	40.3	9.11
B2LU29	Sulfotransferase	0.00	5.71	12	1	2	3	3.544E7	350	40.4	5.73
Q84JZ4	Homoserine dehydrogenase	2.15	2.66	4	1	1	1	1.145E7	376	40.4	6.74
Q9SRT9	UDP-arabinopyranose mutase 1	41.41	21.01	3	5	5	14	9.838E7	357	40.6	5.92
Q39242	Thioredoxin reductase 2	17.16	12.53	2	4	4	5	2.197E7	383	40.6	6.70
Q9LMJ7	At1g07040	7.58	6.47	1	1	1	3	1.979E7	371	40.6	8.78
O22873	BZIP transcription factor	5.05	4.90	2	1	1	1	1.101E7	367	40.6	6.52
Q96533	Alcohol dehydrogenase class-3	20.86	9.76	3	2	2	7	5.887E7	379	40.7	6.95
Q42561	Oleoyl-acyl carrier protein thioesterase 1, chloroplastic	6.90	7.46	2	2	2	4	2.564E7	362	40.8	7.47
Q84W65	SufE-like protein 1, chloroplastic/mitochondrial	8.24	5.93	1	1	1	2	1.276E7	371	40.8	5.64

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9ZUC1	Quinone oxidoreductase-like protein At1g23740, chloroplastic	17.04	10.62	2	4	4	7	5.619E7	386	41.0	8.35
Q8LDG5	Putative uncharacterized protein	2.16	3.54	3	1	1	1	2.631E6	367	41.0	4.96
F4I032	Chorismate synthase	1.88	2.37	3	1	1	1	2.436E7	380	41.0	8.81
Q9SLA8	Enoyl-[acyl-carrier-protein] reductase [NADH], chloroplastic	26.13	7.18	1	2	2	6	8.093E7	390	41.2	9.00
F4J8V9	Actin 2	15.76	17.25	22	1	5	7	1.618E8	371	41.2	5.69
P92966	Serine/arginine-rich splicing factor RS41	2.72	2.81	1	1	1	1	3.318E7	356	41.3	10.32
F4J244	Cystathionine beta-lyase	8.17	7.67	2	2	2	3	2.383E7	378	41.3	8.37
O22791	Putative RNA-binding protein	12.40	5.45	1	1	1	3	1.973E7	404	41.3	6.55
F4HUA0	Elongation factor 1-alpha	19.81	11.56	10	4	4	7	1.554E9	372	41.3	9.23
Q8S4Y1	Acetyl-CoA acetyltransferase, cytosolic 1	18.13	20.10	1	4	4	5	2.747E7	403	41.4	6.89
Q8RU07	Putative malonyl-CoA	4.45	6.36	2	2	2	2	1.864E7	393	41.5	8.56
Q940G9	Periaxin-like protein	2.44	4.59	2	1	1	1	1.278E7	370	41.6	6.18
B3H684	Nucleosome assembly protein 1-like 1	1.94	3.34	2	1	1	1	4.127E7	359	41.6	4.45
P47999	Cysteine synthase, chloroplastic/chromoplastic	4.68	12.50	1	2	2	2	3.193E7	392	41.6	8.02
P53496	Actin-11	35.10	20.69	20	1	7	15	4.104E8	377	41.6	5.39
Q8L5Z1	GDSL esterase/lipase At1g33811	0.00	5.41	1	1	1	1	3.548E7	370	41.7	9.28
P53492	Actin-7	43.05	26.26	21	2	9	18	4.393E8	377	41.7	5.49
F1LIM0	Actin 1	7.37	15.65	17	1	5	5	1.126E8	377	41.7	5.49
Q9FPF0	Protein DJ-1 homolog A	5.32	8.93	2	2	2	3	1.757E7	392	41.8	5.41
O64530	Thiosulfate/3-mercaptopyruvate sulfurtransferase 1, mitochondrial	2.42	7.92	1	1	1	1	8.978E6	379	41.9	6.39
Q9ZVI0	Putative uncharacterized protein At2g38580	1.96	4.77	2	1	1	1		377	41.9	4.73
O04904	Dihydroorotase, mitochondrial	22.52	12.73	1	4	4	8	7.320E7	377	41.9	8.46
P93031	GDP-mannose 4,6 dehydratase 2	2.28	3.49	1	1	1	1	2.641E7	373	41.9	6.11
O81014	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, chloroplastic	2.08	3.39	1	1	1	1	2.629E6	383	42.0	7.02
Q9LQ22	F14M2.18 protein	11.71	11.05	2	2	2	3	4.583E7	389	42.1	4.59
Q84MD8	Bifunctional riboflavin kinase/FMN phosphatase	2.43	2.37	1	1	1	1	7.024E6	379	42.1	6.38
Q8LFV7	Phosphoglycerate kinase	11.18	14.46	3	5	5	10	3.447E8	401	42.1	5.68
Q9STM6	GDSL esterase/lipase At3g48460	1.85	1.84	4	1	1	1	9.117E6	381	42.2	7.74
O23606	NAK like protein kinase	0.00	2.70	2	1	1	1	2.608E7	371	42.2	9.07
Q94KD0	Transcription initiation factor TFIID subunit 15b	4.88	8.06	1	2	2	2	1.710E7	422	42.3	8.50
Q9FPJ8	Polyadenylate-binding protein RBP45A	3.93	5.43	1	1	1	2	2.829E7	387	42.3	6.55
Q9ZU52	Probable fructose-bisphosphate aldolase 3, chloroplastic	12.14	6.91	1	3	3	7	2.343E8	391	42.3	8.09
Q9SN86	Malate dehydrogenase, chloroplastic	0.00	4.71	1	1	1	3	2.655E7	403	42.4	8.51
Q9FM47	RNA-binding (RRM/RBD/RNP motifs) family protein	2.02	6.38	1	1	1	1	6.026E6	423	42.4	6.47
Q9S7E4	Formate dehydrogenase, mitochondrial	1.92	1.82	1	1	1	1	8.724E7	384	42.4	7.50
P46283	Sedoheptulose-1,7-bisphosphatase, chloroplastic	12.97	5.60	1	3	3	6	8.422E7	393	42.4	6.57
P25856	Glyceraldehyde-3-phosphate dehydrogenase GAPA1, chloroplastic	10.84	10.35	3	1	3	4	6.900E7	396	42.5	7.75
Q9SJL8	S-adenosylmethionine synthase 3	31.20	17.44	1	4	6	16	6.558E8	390	42.5	6.18
Q3EAC9	Uncharacterized protein	4.90	2.76	1	1	1	1	9.618E7	579	42.5	12.76
Q9XJ35	ATP-dependent Clp protease proteolytic subunit-related protein 1, chloroplastic	0.00	3.10	1	1	1	1	1.009E7	387	42.6	8.63
Q9S9P8	Ferredoxin--NADP reductase, root isozyme 2, chloroplastic	5.88	5.76	1	2	2	4	2.674E7	382	42.8	8.60
Q9LUT2	S-adenosylmethionine synthase 4	34.59	18.07	1	2	6	18	7.627E8	393	42.8	5.81
Q9LNC6	At1g06210/F9P14_4	4.16	4.44	1	1	1	1	5.194E7	383	42.8	4.78
F4ISI7	Nucleosome assembly protein 12	23.47	14.25	3	5	5	9	3.616E7	372	42.8	4.40
Q8L637	Putative uncharacterized protein At3g21140	2.65	7.75	2	1	1	1	8.325E6	387	42.8	6.93

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9SJU4	Probable fructose-bisphosphate aldolase 1, chloroplastic	2.01	6.52	4	2	2	2	2.304E6	399	42.9	6.58
Q944G9	Probable fructose-bisphosphate aldolase 2, chloroplastic	13.02	16.58	2	4	4	5	2.581E7	398	43.0	7.24
P23686	S-adenosylmethionine synthase 1	44.71	19.34	1	2	8	23	9.387E8	393	43.1	5.82
P49077	Aspartate carbamoyltransferase, chloroplastic	0.00	1.54	1	1	1	1	5.166E7	390	43.1	6.60
Q0WLJ0	Peroxisomal-3-keto-acyl-CoA thiolase 1	5.40	10.39	2	2	2	4	1.176E7	414	43.1	6.79
Q8LEF4	3-isopropylmalate dehydrogenase	5.01	6.44	5	1	2	2	1.437E7	404	43.3	6.06
Q9S850	Sulfite oxidase	13.87	14.25	2	4	4	6	5.024E7	393	43.3	8.68
Q9SU13	Fasciclin-like arabinogalactan protein 2	4.34	10.67	1	3	3	3	6.986E7	403	43.4	6.62
F4K7E0	Nuclear transport factor 2 and RNA recognition motif domain-containing protein	1.62	5.37	2	2	2	2	9.519E7	391	43.5	4.97
Q9LTU9	Genomic DNA, chromosome 3, P1 clone: MDC11	0.00	3.96	3	1	1	1	1.644E7	379	43.5	8.66
Q9FKA5	Uncharacterized protein At5g39570	53.02	42.26	1	8	8	26	1.109E9	381	43.5	4.78
Q8L733	Putative transferase At4g12130, mitochondrial	5.33	7.89	1	2	2	3	7.196E6	393	43.5	6.73
Q9FIQ0	Probable ADP-ribosylation factor GTPase-activating protein AGD9	9.32	3.98	1	1	1	2	2.228E7	402	43.5	8.32
O22886	Uroporphyrinogen decarboxylase 2, chloroplastic	9.21	6.35	2	2	2	3	5.461E7	394	43.6	8.21
Q9C969	At1g80360	4.67	5.58	1	1	1	2	3.985E7	394	43.7	6.37
Q9LR75	Coproporphyrinogen-III oxidase 1, chloroplastic	9.19	8.29	2	2	2	6	1.135E8	386	43.8	6.70
Q9LYA9	Chloroplast stem-loop binding protein of 41 kDa a, chloroplastic	10.98	7.64	1	3	3	3	3.794E7	406	43.9	8.43
Q8GYA6	26S proteasome non-ATPase regulatory subunit 13 homolog B	7.57	2.07	2	1	1	4	1.532E7	386	44.0	5.24
Q93Z70	Probable N-acetyl-gamma-glutamyl-phosphate reductase, chloroplastic	19.23	8.48	1	2	2	6	3.161E7	401	44.1	8.29
P46645	Aspartate aminotransferase, cytoplasmic isozyme 1	9.23	10.12	2	2	3	3	2.807E7	405	44.2	7.28
Q93Y35	26S proteasome non-ATPase regulatory subunit 6 homolog	0.00	7.75	1	2	2	2	4.533E7	387	44.3	5.99
O65396	Aminomethyltransferase, mitochondrial	18.20	13.48	1	7	7	14	1.574E8	408	44.4	8.37
Q8S948	Solanesyl diphosphate synthase 1	1.89	1.72	1	1	1	1	1.835E8	406	44.4	5.22
Q9SA73	Obg-like ATPase 1	0.00	3.55	1	1	1	2	3.760E7	394	44.4	6.81
Q2V3X4	60S ribosomal protein L4-1	27.55	17.04	4	2	6	12	2.185E8	405	44.5	10.37
F4KDU5	60S ribosomal protein L4-2	31.01	17.73	4	2	6	12	1.531E8	406	44.6	10.35
Q38946	Glutamate dehydrogenase 2	2.40	4.62	2	2	2	2	3.875E7	411	44.7	6.54
Q5E924	Glyceraldehyde-3-phosphate dehydrogenase GAPCP2, chloroplastic	2.96	3.33	2	1	1	1	3.163E7	420	44.8	8.62
Q9FM65	Fasciclin-like arabinogalactan protein 1	6.71	3.30	1	1	1	2	5.918E8	424	44.8	6.74
Q9SV55	AFP homolog 2	8.10	8.00	1	1	1	2	3.441E7	425	44.9	5.78
Q8L5U0	COP9 signalosome complex subunit 4	1.65	2.52	1	1	1	1	7.550E6	397	44.9	4.98
Q93W34	Polyadenylate-binding protein RBP45C	13.57	9.40	1	2	2	4	3.700E7	415	44.9	5.90
Q7Y175	Plant UBX domain-containing protein 5	2.52	2.85	1	1	1	1	2.225E7	421	45.0	5.62
Q9LF04	Acyl-[acyl-carrier-protein] desaturase 1, chloroplastic	7.40	12.18	1	3	3	5	1.841E7	394	45.0	6.55
F4I577	Monodehydroascorbate reductase	11.42	16.59	4	5	5	5	3.881E7	416	45.0	8.59
O49429	Multiple organellar RNA editing factor 1, mitochondrial	0.00	5.49	1	1	1	2	4.654E7	419	45.1	7.84
P25851	Fructose-1,6-bisphosphatase, chloroplastic	10.58	12.95	1	3	3	8	9.932E7	417	45.1	5.40
O82662	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	10.02	11.64	3	5	5	7	1.711E8	421	45.3	6.71
Q9FIY1	Patatin-like protein 4	0.00	2.99	1	1	1	2	3.405E7	401	45.4	5.47
Q9SGT3	Obg-like ATPase 1	1.89	2.86	2	1	1	1	5.728E6	419	45.4	5.81
Q9XI49	F9L1.14 protein	2.35	3.54	5	1	1	1	2.760E7	395	45.4	5.39
Q8H183	Beta-ureidopropionase	9.14	9.07	1	2	2	3	2.089E7	408	45.5	6.35
P49063	Exopolygalacturonase clone GBGA483	2.25	2.93	5	1	1	1	3.600E7	444	45.6	8.31
F4IS32	Acyl-[acyl-carrier-protein] desaturase	2.02	2.24	2	1	1	3	1.109E8	401	45.6	6.64
C0Z2P8	AT1G80460 protein	2.25	8.98	3	2	2	2	4.269E7	423	45.7	6.93

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9FT97	Alpha-galactosidase 1	0.00	2.68	1	1	1	1	4.537E6	410	45.7	6.67
Q9SRZ6	Cytosolic isocitrate dehydrogenase [NADP]	17.89	16.59	5	7	7	10	1.414E8	410	45.7	6.57
Q94AT1	Probable protein phosphatase 2C 76	7.90	8.10	1	2	2	5	9.245E6	420	45.8	5.10
Q9LFX8	Putative uncharacterized protein At1g27090	9.49	7.14	1	1	1	2	2.198E7	420	46.0	5.39
Q42593	L-ascorbate peroxidase T, chloroplastic	1.86	2.35	1	1	2	3	1.054E8	426	46.1	7.28
Q8RW90	At1g12050/F12F1_8	5.26	2.38	1	1	1	2	1.824E8	421	46.1	5.55
Q8VYA0	Aspartate--tRNA ligase-like protein	9.41	4.20	2	1	1	2	1.914E7	405	46.1	7.15
O82264	NPL4-like protein 2	7.98	4.12	1	1	1	3	3.604E7	413	46.1	5.15
Q9T0K7	3-hydroxyisobutyryl-CoA hydrolase-like protein 3, mitochondrial	11.25	8.08	1	3	3	5	2.391E7	421	46.2	6.67
Q93ZB6	Uroporphyrinogen decarboxylase 1, chloroplastic	7.02	3.83	1	1	1	3	8.555E7	418	46.2	7.11
Q9FVT2	Probable elongation factor 1-gamma 2	39.09	13.56	4	6	7	15	9.290E8	413	46.4	5.71
Q9LVY1	Tyrosine aminotransferase	6.10	4.05	1	1	1	2	7.160E7	420	46.4	5.02
Q9LFA3	Probable monodehydroascorbate reductase, cytoplasmic isoform 3	45.31	20.05	2	6	6	15	2.419E8	434	46.5	6.83
Q9FN02	Serine/threonine-protein phosphatase 7	0.00	2.42	2	1	1	1	8.004E6	413	46.6	6.02
Q93YR3	FAM10 family protein At4g22670	0.00	3.40	1	1	1	1	2.477E7	441	46.6	4.93
O04487	Probable elongation factor 1-gamma 1	12.64	10.63	4	3	4	7	7.403E7	414	46.6	5.48
Q8GWP5	Riboflavin biosynthesis protein PYRD, chloroplastic	4.62	2.82	1	1	1	2	2.287E7	426	46.6	6.46
Q9FX53	Trihelix transcription factor GT-1	1.71	1.97	1	1	1	1	1.039E7	406	46.6	6.87
Q9SGY2	ATP-citrate synthase alpha chain protein 1	1.67	3.78	4	2	2	2	2.207E7	423	46.6	5.50
Q9SJ62	At2g35880/F11F19.21	1.83	2.78	1	1	1	1	2.332E7	432	46.7	9.70
Q9LVI9	Dihydropyrimidine dehydrogenase (NADP(+)), chloroplastic	8.99	8.22	1	2	2	5	7.652E7	426	46.8	6.80
F4J440	Alpha/beta-Hydrolases superfamily protein	3.85	3.55	3	1	1	1	1.015E7	423	46.9	5.10
Q9ZV76	Putative uncharacterized protein At2g07180	0.00	1.89	3	1	1	1	8.091E6	423	47.0	8.53
C0Z2C0	AT1G56340 protein	38.90	26.28	8	7	9	15	2.161E8	411	47.0	4.67
Q9FN03	Ultraviolet-B receptor UVR8	7.27	5.91	1	1	1	2	4.179E7	440	47.1	5.74
O24457	Pyruvate dehydrogenase E1 component subunit alpha-3, chloroplastic	0.00	2.10	1	1	1	1	1.120E8	428	47.1	7.49
B9DFF8	AT4G14960 protein	24.51	12.65	6	3	3	9	2.567E8	427	47.2	8.09
Q39048	Protein ECERIFERUM 1	2.97	4.04	1	1	1	1	8.128E6	421	47.2	5.60
Q96255	Phosphoserine aminotransferase 1, chloroplastic	0.00	1.86	1	1	1	1	1.506E7	430	47.3	8.06
Q8LBB2	SNF1-related protein kinase regulatory subunit gamma-1	5.57	4.01	1	1	1	2	1.524E7	424	47.4	5.30
Q9SEI2	26S protease regulatory subunit 6A homolog A	9.00	8.96	2	3	3	4	3.265E7	424	47.5	5.03
P25857	Glyceraldehyde-3-phosphate dehydrogenase GAPB, chloroplastic	11.21	8.95	1	2	3	5	9.039E7	447	47.6	6.80
Q9LYR4	Aldolase-type TIM barrel family protein	25.78	12.56	1	5	5	9	5.131E7	438	47.7	6.43
P25696	Bifunctional enolase 2/transcriptional activator	60.14	34.23	3	12	12	26	5.977E8	444	47.7	5.77
O48773	Protein disulfide-isomerase 2-3	3.53	7.05	1	2	2	3	3.556E7	440	47.7	5.85
P46643	Aspartate aminotransferase, mitochondrial	9.40	14.42	1	4	4	6	2.783E7	430	47.7	8.19
Q9SSB5	26S protease regulatory subunit 7 homolog A	7.33	14.32	1	4	4	6	9.076E7	426	47.8	6.65
Q38858	Calreticulin-2	14.30	16.04	2	4	6	10	1.100E8	424	48.1	4.53
Q9FHP0	BSD domain-containing protein	0.00	4.63	1	1	1	2	4.136E7	432	48.2	4.40
Q9C5C4	Acetylornithine deacetylase	6.87	5.00	2	1	1	2	1.359E7	440	48.2	5.35
Q9LKR8	Rubisco accumulation factor 1, chloroplastic	0.00	1.84	1	1	1	1	4.358E7	434	48.2	5.07
Q8H135	ATP synthase subunit beta (Fragment)	20.24	11.43	4	5	6	8	9.550E7	446	48.2	5.63
F4I3B3	Polyadenylate-binding protein RBP47A	2.20	2.92	1	1	1	1	4.371E6	445	48.5	6.00
Q56WD9	3-ketoacyl-CoA thiolase 2, peroxisomal	36.04	15.37	3	6	6	13	1.296E8	462	48.5	8.34
Q9ZUR7	Arginine biosynthesis bifunctional protein ArgJ, chloroplastic	4.78	5.34	1	2	2	2	1.591E7	468	48.7	6.65
Q9M8M7	Acetylornithine aminotransferase, chloroplastic/mitochondrial	2.61	3.50	1	1	1	1	1.175E7	457	48.8	6.80

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9SZP8	Glycine-rich protein	2.90	6.19	2	2	2	3	1.111E7	452	49.0	5.14
O48832	Probable senescence related protein	2.04	2.43	2	1	1	1	1.460E7	452	49.0	5.41
B3H778	Argininosuccinate synthase	0.00	4.89	2	1	1	1	2.332E7	450	49.0	7.72
Q56Y85	Methionine aminopeptidase 2B	4.97	4.33	1	1	1	1	1.919E7	439	49.0	5.69
F4JWW6	Omega-hydroxypalmitate O-feruloyl transferase	2.52	3.40	2	1	1	1	9.995E6	441	49.2	5.80
F4JTH0	Aspartate aminotransferase	9.82	15.18	5	5	6	7	7.779E7	448	49.3	7.87
Q9SL67	26S proteasome regulatory subunit 4 homolog B	4.53	2.26	2	2	2	2	1.197E8	443	49.3	6.10
Q8S9J9	At1g14000/F7A19_9	5.92	5.71	2	1	1	2	5.500E7	438	49.3	7.94
Q9ZT91	Elongation factor Tu, mitochondrial	9.88	2.20	1	1	1	4	5.860E8	454	49.4	6.70
Q8RWB6	Putative serine protease-like protein	3.22	2.76	1	1	1	1	1.698E7	434	49.4	5.00
F4K1Y4	Nuclear transport factor 2 and RNA recognition motif domain-containing protein	2.00	2.18	3	1	1	1	1.757E8	459	49.4	5.96
Q6NPN3	Zinc finger CCCH domain-containing protein 58	0.00	2.15	1	1	1	1		465	49.4	6.86
O24653	Guanosine nucleotide diphosphate dissociation inhibitor 2	0.00	3.83	3	1	1	1	4.598E7	444	49.5	5.74
Q9M354	Probable ADP-ribosylation factor GTPase-activating protein AGD6	2.18	3.70	1	1	1	3	1.566E7	459	49.7	7.24
Q9LTR9	Glycylpeptide N-tetradecanoyltransferase 1	3.14	2.76	1	1	1	1	1.212E8	434	49.8	6.80
A7KNE3	ATP synthase subunit alpha (Fragment)	10.55	4.54	5	2	2	5	8.426E7	463	49.8	6.30
Q9SND9	Uncharacterized acetyltransferase At3g50280	0.00	3.84	1	1	1	1	4.342E6	443	49.9	6.71
Q8H107	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex 2, mitochondrial	24.19	6.25	1	1	1	6	1.296E8	464	50.0	9.09
F4JWN4	Uncharacterized protein	0.00	2.28	4	1	1	1	6.502E6	439	50.0	5.39
Q84JH2	Nuclear transport factor 2 and RNA recognition motif domain-containing protein	3.14	3.71	2	1	1	1	1.695E7	458	50.0	5.41
Q9LD57	Phosphoglycerate kinase 1, chloroplastic	27.08	22.25	3	9	9	18	3.330E8	481	50.1	6.24
Q9FLQ4	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex 1, mitochondrial	22.95	10.78	1	4	4	7	4.149E7	464	50.1	9.14
Q42522	Glutamate-1-semialdehyde 2,1-aminomutase 2, chloroplastic	5.65	9.32	2	3	3	3	1.088E8	472	50.1	7.39
Q9SR19	Rubisco accumulation factor 2, chloroplastic	4.19	1.78	1	1	1	3	3.730E7	449	50.2	5.96
O80724	Expressed protein	0.00	2.66	2	1	1	1	2.569E7	451	50.2	7.46
Q9C5U8	Histidinol dehydrogenase, chloroplastic	13.60	6.22	1	3	3	5	1.624E7	466	50.3	6.16
Q8LBD2	Putative myrosinase-binding protein	0.00	3.06	1	1	1	1	1.439E7	458	50.3	5.24
Q42404	U1 small nuclear ribonucleoprotein 70 kDa	7.23	7.03	1	2	2	3	5.620E7	427	50.4	9.39
Q93ZN9	LL-diaminopimelate aminotransferase, chloroplastic	34.03	5.21	1	2	2	7	6.149E7	461	50.4	7.34
O23153	Putative amidase	0.00	3.28	5	1	1	1	7.484E6	457	50.4	5.99
P52410	3-oxoacyl-[acyl-carrier-protein] synthase I, chloroplastic	6.41	2.75	2	1	1	2	1.213E7	473	50.4	8.06
Q9LK21	Probable carboxylesterase 11	1.80	3.04	1	1	1	1	2.215E7	460	50.5	7.87
A0A097PS65	ARM repeat superfamily protein (Fragment)	2.71	3.88	5	1	1	1	5.624E6	464	50.5	5.81
F4INR3	Aminopeptidase family protein	35.94	21.07	4	3	6	12	1.922E8	484	50.6	6.05
O82533	Cell division protein FtsZ homolog 2-1, chloroplastic	2.75	2.30	1	1	1	1	1.561E7	478	50.7	5.71
Q56YW9	Tubulin beta-2 chain	25.71	16.89	11	6	6	10	5.228E7	450	50.7	4.81
O64490	F20D22.6 protein	0.00	4.52	2	1	1	1	4.954E6	465	50.8	8.76
O65902	Cyclase-associated protein 1	0.00	3.99	1	2	2	2	3.371E7	476	50.9	6.65
Q9SIE1	Bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase	2.20	3.16	1	1	1	1	2.381E6	475	51.0	7.90
P54873	Hydroxymethylglutaryl-CoA synthase	0.00	4.77	1	1	1	1	4.277E7	461	51.1	6.40
Q9FMH8	Cysteine protease component of protease-inhibitor complex	4.38	4.10	2	1	2	2	4.490E7	463	51.2	6.21

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A075M556	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (Fragment)	303.52	51.61	10	29	29	127	9.685E9	465	51.3	6.38
Q9LES2	UBP1-associated protein 2A	4.81	3.14	1	1	1	1	5.593E6	478	51.4	5.16
Q9ZRA2	Homogentisate 1,2-dioxygenase	3.83	3.04	1	1	1	1	4.231E6	461	51.4	6.18
Q9LIK9	ATP sulfurylase 1, chloroplastic	4.63	3.67	4	2	2	2	3.565E7	463	51.4	6.81
Q9LZM1	Oxysterol-binding protein-related protein 3A	1.69	6.18	2	1	2	2	1.293E7	453	51.5	5.24
Q39230	Serine--tRNA ligase	3.09	3.55	1	1	1	1	3.765E7	451	51.6	6.67
O23254	Serine hydroxymethyltransferase 4	43.83	33.33	2	10	10	16	2.901E8	471	51.7	7.23
Q9M9P3	Probable UTP--glucose-1-phosphate uridylyltransferase 2	34.43	20.47	1	4	9	17	2.863E8	469	51.7	6.13
Q940Z5	Phenolic glucoside malonyltransferase 1	0.00	5.54	1	2	2	2	6.295E6	469	51.7	5.55
Q9ZVT6	BSD domain-containing protein	1.98	2.34	1	1	1	1	6.485E6	470	51.9	4.36
Q94AL9	Alanine--glyoxylate aminotransferase 2 homolog 2, mitochondrial	1.63	1.68	2	1	1	1	1.493E7	477	51.9	7.50
P57751	UTP--glucose-1-phosphate uridylyltransferase 1	41.48	21.06	2	4	9	18	2.799E8	470	51.9	6.01
Q940M2	Alanine--glyoxylate aminotransferase 2 homolog 1, mitochondrial	0.00	10.71	1	3	3	4	5.517E7	476	51.9	7.85
Q9XFS9	1-deoxy-D-xylulose 5-phosphate reductoisomerase, chloroplastic	2.83	2.94	2	1	1	1	7.067E6	477	51.9	7.05
Q93Y40	Oxysterol-binding protein-related protein 3C	9.45	8.32	1	2	3	4	1.682E7	457	51.9	5.30
P10896	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic	55.06	13.71	5	5	5	15	1.577E8	474	51.9	6.15
Q8VYQ8	Putative uncharacterized protein At1g35470	2.09	2.37	3	1	1	1	7.969E6	465	52.1	6.23
Q683B7	Prolyl carboxypeptidase like protein (Fragment)	2.80	5.19	5	1	1	3	1.085E8	462	52.1	6.52
Q9M052	UDP-glycosyltransferase 76F1	0.00	4.57	1	1	1	1	8.717E6	460	52.1	6.07
Q8LF32	Aspartyl aminopeptidase-like protein	0.00	2.10	2	1	1	1	1.770E7	477	52.5	6.77
Q8LEZ6	Adenylosuccinate synthetase, chloroplastic	3.54	4.49	2	1	1	1	1.474E7	490	52.9	7.14
B9DHU0	AT1G20620 protein (Fragment)	39.48	20.31	10	7	8	17	1.294E8	458	53.0	7.46
B9DI07	AT1G74040 protein (Fragment)	2.98	2.04	3	1	1	1	3.721E7	491	53.0	5.73
Q1WIQ6	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	14.31	5.85	2	3	3	6	1.590E8	496	53.0	6.64
B9DFH4	AT2G47510 protein	17.08	12.60	5	6	6	7	2.191E8	492	53.0	7.88
Q9LF33	UDP-glucose 6-dehydrogenase 3	3.83	5.21	2	2	2	2	4.995E7	480	53.1	6.04
Q9LR30	Glutamate--glyoxylate aminotransferase 1	8.23	3.74	2	1	1	2	2.368E7	481	53.3	6.89
Q9FFR3	6-phosphogluconate dehydrogenase, decarboxylating 2, chloroplastic	25.49	13.55	1	2	5	9	4.890E7	487	53.3	5.80
Q9SH69	6-phosphogluconate dehydrogenase, decarboxylating 1, chloroplastic	37.57	15.81	1	3	6	13	5.267E7	487	53.3	5.45
A8MQR6	Methylmalonate-semialdehyde dehydrogenase [acylating]	33.41	5.42	3	3	3	8	3.057E7	498	53.4	6.00
Q9FNN5	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	2.56	7.82	1	2	2	3	4.683E7	486	53.4	8.16
F4JMS5	Glutamate-cysteine ligase	2.10	2.94	2	1	1	1	2.358E7	477	53.4	7.66
Q9FF86	BAHD acyltransferase DCR	5.29	8.26	1	3	3	4	1.065E7	484	53.5	5.49
Q9FWA3	6-phosphogluconate dehydrogenase, decarboxylating 3	16.17	10.91	1	3	4	8	9.220E7	486	53.5	7.42
Q9LEY1	Serine carboxypeptidase-like 35	2.19	3.75	1	1	1	1		480	53.6	7.90
P48641	Glutathione reductase, cytosolic	12.61	7.62	1	3	3	5	7.439E7	499	53.8	6.80
P19366	ATP synthase subunit beta, chloroplastic	26.42	17.87	3	4	5	8	4.325E7	498	53.9	5.50
F4HXD2	Aldehyde dehydrogenase 10A8	0.00	1.81	2	1	1	1	2.283E7	496	53.9	5.35
Q9M5K2	Dihydrolipoyl dehydrogenase 2, mitochondrial	9.61	11.83	1	3	4	4	3.301E7	507	54.0	7.03
Q9M5K3	Dihydrolipoyl dehydrogenase 1, mitochondrial	12.53	13.41	2	4	5	6	8.306E7	507	54.0	7.40
O23264	Selenium-binding protein 1	1.99	3.47	2	2	2	2	2.334E7	490	54.0	5.67
P11574	V-type proton ATPase subunit B1	1.72	1.65	4	1	1	1	4.423E7	486	54.1	5.10
F4HZN9	Protein disulfide-isomerase	48.12	18.28	2	5	6	17	4.345E8	487	54.1	5.08
Q9ZQK6	Putative uncharacterized protein At2g07360	18.58	5.62	3	2	2	8	3.396E7	498	54.4	4.64
Q94CD8	Glucan endo-1,3-beta-glucosidase 4	8.07	4.36	1	1	1	3	2.345E7	505	54.4	6.35

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9LMF1	UDP-glycosyltransferase 85A3	3.80	3.69	1	2	2	3	2.977E7	488	54.6	5.85
Q9SU83	Alkaline-phosphatase-like protein	3.66	3.63	5	2	2	3	5.056E7	496	54.6	5.88
Q9STS1	Betaine aldehyde dehydrogenase 2, mitochondrial	2.30	3.98	1	1	1	1	2.129E7	503	54.9	5.52
Q9FNN1	Pyruvate kinase	13.35	6.86	7	5	5	6	8.385E7	510	54.9	6.29
Q949Y3	Bifunctional purple acid phosphatase 26	10.29	8.21	1	2	2	5	5.027E7	475	55.0	7.31
Q94CE5	Gamma-aminobutyrate transaminase POP2, mitochondrial	16.89	8.93	1	4	4	5	9.920E7	504	55.2	7.94
P56757	ATP synthase subunit alpha, chloroplastic	8.95	4.93	2	3	3	5	5.299E7	507	55.3	5.25
Q9T0D8	Putative uncharacterized protein AT4g11710	0.00	1.90	1	1	1	1	1.084E8	473	55.5	9.33
P05466	3-phosphoshikimate 1-carboxyvinyltransferase, chloroplastic	5.49	4.81	4	2	2	2	4.807E7	520	55.7	6.70
Q9FFW8	Tryptophan synthase beta chain	0.00	1.78	1	1	1	1	7.681E6	506	55.7	7.24
O22607	WD-40 repeat-containing protein MSI4	2.82	2.56	1	1	1	1	5.578E7	507	55.7	6.23
F4JVC5	Probable phenylalanine--tRNA ligase alpha subunit	2.52	3.09	2	1	1	2	3.405E7	485	55.7	8.34
Q8VYJ6	At2g30880/F7F1.9	3.33	3.17	2	1	1	1	1.786E7	504	56.0	5.87
P25853	Beta-amylase 5	13.15	11.45	1	3	3	9	9.459E7	498	56.0	5.36
Q9FIW4	Beta-glucosidase 42	5.97	2.65	1	1	1	2	2.145E7	490	56.0	5.63
Q56Z59	Patellin-3	2.24	2.24	1	1	1	1	4.271E7	490	56.1	5.30
Q9SRG3	Protein disulfide isomerase-like 1-2	15.26	12.80	1	5	6	9	2.862E8	508	56.3	5.00
Q9FGF3	At5g64430	3.52	3.51	1	1	1	1	4.935E6	513	56.4	5.94
P52420	Phosphoribosylamine-glycine ligase, chloroplastic	0.00	2.26	1	1	1	1	8.546E6	532	56.4	5.67
Q9SJX2	Putative kinesin light chain	2.67	2.35	32	2	2	2	4.637E7	510	56.5	5.64
C4PW06	Inositol-3-phosphate synthase	14.94	6.85	9	3	3	9	4.196E7	511	56.5	5.66
Q9FFB0	Serine carboxypeptidase-like 47	0.00	2.97	1	1	1	1	7.329E6	505	56.5	7.39
Q9LTM8	Cellulase (Glycosyl hydrolase family 5) protein	11.29	4.92	1	2	2	4	6.211E7	508	56.5	6.14
Q9SAK4	Succinate-semialdehyde dehydrogenase, mitochondrial	4.58	3.41	1	2	2	2	3.624E7	528	56.5	6.92
Q96528	Catalase-1	21.14	15.04	1	5	6	10	9.186E7	492	56.7	7.42
Q38821	Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	6.21	3.90	38	2	2	3	1.494E7	513	56.9	5.72
Q9ZW74	Plant UBX domain-containing protein 11	0.00	3.95	1	1	1	1	1.512E7	531	57.1	5.53
Q8H7E1	Amine oxidase	2.46	11.83	5	5	5	5	8.282E7	507	57.1	5.55
Q9FM80	Mitochondrial transcription termination factor family protein	1.64	1.21	1	1	1	1	4.546E8	496	57.2	8.79
F4KCE5	Heat shock cognate protein 70-1	113.82	25.53	5	5	11	34	6.917E8	521	57.2	5.10
Q940P8	T-complex protein 1 subunit beta	2.04	1.90	1	1	1	1	2.260E7	527	57.2	5.87
Q94C74	Serine hydroxymethyltransferase 2, mitochondrial	5.49	8.51	1	1	3	3	4.663E7	517	57.3	8.75
Q9SZJ5	Serine hydroxymethyltransferase 1, mitochondrial	43.65	30.17	3	7	9	19	1.361E8	517	57.4	8.13
P55230	Glucose-1-phosphate adenylyltransferase large subunit 2, chloroplastic	0.00	5.02	1	1	1	3		518	57.4	7.47
A0A068LMZ4	DM3Hh0	0.00	2.33	2	1	1	1	1.056E7	515	57.5	7.31
F4J5F4	Plant transposase (Ptta/En/Spm family)	7.63	3.14	2	1	1	2	5.843E6	510	57.6	8.59
P55229	Glucose-1-phosphate adenylyltransferase large subunit 1, chloroplastic	5.36	3.83	1	2	2	2	6.112E7	522	57.6	7.91
Q93Y22	Coatomer subunit delta	8.00	5.88	1	1	1	2	2.129E7	527	57.7	5.85
Q9LIN5	Eukaryotic initiation factor 4B	25.72	12.22	3	4	4	8	4.390E7	532	57.7	8.13
Q9S7B5	Threonine synthase 1, chloroplastic	12.54	9.13	1	4	4	5	5.076E7	526	57.7	7.42
Q94JQ3	Serine hydroxymethyltransferase 3, chloroplastic	8.75	8.13	1	4	4	6	8.503E7	529	57.9	8.88
P55231	Glucose-1-phosphate adenylyltransferase large subunit 3, chloroplastic	3.24	2.69	1	1	1	1	6.503E7	521	58.0	8.41
Q9SII8	Ubiquitin domain-containing protein DSK2b	1.98	5.44	2	2	2	2	1.013E7	551	58.0	4.93
Q9LN93	F5O11.31	15.18	5.57	4	2	2	5	7.066E7	539	58.0	8.59
Q0WWW1	Glycyl-tRNA synthetase (Fragment)	9.88	5.20	2	2	2	3	6.238E7	519	58.1	5.99

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
O04983	Biotin carboxylase, chloroplastic	30.31	19.55	3	7	7	11	7.966E7	537	58.3	7.25
B9DI42	AT1G12800 protein (Fragment)	0.00	2.29	3	1	1	1	1.054E7	525	58.4	4.82
Q9FKG3	ATP-dependent 6-phosphofructokinase 4, chloroplastic	5.94	2.83	1	1	1	2	1.012E7	530	58.4	8.24
Q5M729	Dihydrolipoyllysine-residue acetyltransferase component 3 of pyruvate dehydrogenase complex, mitochondrial	0.00	1.86	1	1	1	1	3.816E7	539	58.4	7.84
Q8RWN9	Dihydrolipoyllysine-residue acetyltransferase component 2 of pyruvate dehydrogenase complex, mitochondrial	10.18	12.62	1	5	5	11	8.803E7	539	58.4	7.65
A8MS69	4-coumarate--CoA ligase 1	12.55	7.61	8	2	3	7	1.729E8	539	58.5	5.10
Q9SU63	Aldehyde dehydrogenase family 2 member B4, mitochondrial	8.75	4.65	1	3	3	5	1.193E8	538	58.6	7.46
Q96321	Importin subunit alpha-1	6.02	5.83	2	2	2	2	6.332E6	532	58.6	5.27
Q9LFV2	Putative uncharacterized protein F14F8_120	0.00	1.58	3	1	1	1	4.998E7	505	58.6	9.00
Q94K05	T-complex protein 1 subunit theta	3.83	4.01	1	1	1	1	2.967E7	549	58.9	5.35
Q9M888	T-complex protein 1 subunit zeta 1	2.23	3.93	2	2	2	2	1.818E7	535	58.9	6.21
Q56ZK2	Putative uncharacterized protein At1g01980 (Fragment)	2.10	2.07	4	1	1	1	5.086E7	532	59.0	8.84
Q94BN5	Putative uncharacterized protein At5g50900	2.56	3.06	2	1	1	1	1.798E7	555	59.2	7.81
Q42479	Calcium-dependent protein kinase 3	2.18	2.84	1	2	2	3	1.653E7	529	59.3	6.37
Q9LTX3	Pyridoxine/pyridoxamine 5'-phosphate oxidase 1, chloroplastic	21.30	15.47	1	6	6	11	7.824E7	530	59.3	7.36
O04450	T-complex protein 1 subunit epsilon	0.00	3.74	1	1	1	1	2.679E7	535	59.3	5.66
Q9LDV4	Alanine aminotransferase 2, mitochondrial	6.93	8.89	1	2	3	5	2.134E7	540	59.5	6.38
Q84WM2	At3g15730/MSJ11_13	6.72	5.35	2	2	2	6	8.017E7	523	59.5	5.30
Q9SF16	T-complex protein 1 subunit eta	2.42	3.77	1	2	2	2	2.031E7	557	59.7	6.39
F4I7I0	Alanine aminotransferase 1, mitochondrial	23.66	11.79	1	7	8	15	2.073E8	543	59.8	6.43
O64858	At2g44230/F4I1.4	7.11	3.69	1	2	2	3	1.015E7	542	59.9	5.91
F4HQR5	Protein phosphatase 2A subunit A3	0.00	2.42	2	1	1	1	1.121E7	537	60.0	5.22
F4JLP5	Dihydrolipoyl dehydrogenase 2, chloroplastic	11.53	20.28	1	3	7	8	1.128E8	567	60.1	7.61
Q84WV1	T-complex protein 1 subunit gamma	2.26	2.16	1	1	1	1	1.438E7	555	60.3	5.77
Q38868	Calcium-dependent protein kinase 9	2.46	2.40	1	1	1	1	2.314E7	541	60.3	6.18
Q9LXD9	Probable pectinesterase/pectinesterase inhibitor 51	0.00	1.09	1	1	1	1	2.393E7	551	60.4	6.37
F4JJU5	Glycosyl hydrolase family protein	2.08	2.08	2	1	1	1	1.088E7	529	60.4	6.57
Q93ZM7	Chaperonin CPN60-like 2, mitochondrial	1.87	2.10	1	2	2	2	2.960E7	572	60.4	6.10
Q9SA79	T5I8.9 protein	0.00	3.45	2	1	1	1	1.578E7	522	60.5	9.17
Q6TBX7	Carotene epsilon-monooxygenase, chloroplastic	0.00	1.67	1	1	1	1	4.847E6	539	60.5	6.39
O04499	2,3-bisphosphoglycerate-independent phosphoglycerate mutase 1	4.21	13.64	4	2	5	8	4.856E7	557	60.5	5.53
Q9LPK6	Probable acyl-activating enzyme 9	1.85	2.18	1	1	1	1	8.738E6	550	60.6	6.98
Q9FYR6	At5g52520	1.88	2.03	1	1	1	1	4.774E7	543	60.7	7.02
Q9M1R2	Class II aaRS and biotin synthetases superfamily protein	7.04	4.91	1	2	2	3	1.172E8	530	60.7	6.40
Q9M9K1	Probable 2,3-bisphosphoglycerate-independent phosphoglycerate mutase 2	12.98	20.00	2	4	7	11	9.708E7	560	60.7	5.85
B1GUZ2	Cinnamyl alcohol dehydrogenase	7.42	5.22	8	1	2	6	1.630E8	556	60.8	5.81
P42770	Glutathione reductase, chloroplastic	12.70	7.61	2	3	3	7	8.921E7	565	60.8	7.87
P37702	Myrosinase 1	19.40	17.38	1	6	7	12	1.368E8	541	61.1	5.92
A7WM73	Beta-hexosaminidase 1	3.83	2.22	1	1	1	2	2.621E7	541	61.2	6.27
P29197	Chaperonin CPN60, mitochondrial	29.53	12.82	1	4	8	11	4.502E7	577	61.2	5.78
Q949W8	Putative xylulose kinase	17.22	5.02	3	2	2	6	7.972E6	558	61.3	5.53
O64745	At2g34810	2.47	2.78	1	1	1	1	6.224E7	540	61.3	9.61
Q944P7	Leucine aminopeptidase 3, chloroplastic	18.51	10.63	5	5	8	9	1.514E8	583	61.3	7.08

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q38931	Peptidyl-prolyl cis-trans isomerase FKBP62	14.05	13.43	3	7	7	9	1.675E8	551	61.4	5.31
Q8W4M5	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta 1	10.90	6.71	2	3	3	5	1.303E8	566	61.4	6.10
Q9LF46	2-hydroxyacyl-CoA lyase	1.61	4.90	1	2	2	2	6.720E6	572	61.4	6.05
F4IVR2	Heat shock protein 60-2	32.31	13.79	2	5	9	12	4.807E7	580	61.4	6.33
Q8VZC3	Delta-1-pyrroline-5-carboxylate dehydrogenase 12A1, mitochondrial	11.86	5.94	1	2	2	4	6.356E7	556	61.7	6.73
P21238	Chaperonin 60 subunit alpha 1, chloroplastic	41.99	31.06	2	10	11	17	6.796E7	586	62.0	5.19
F4JPR7	Plant UBX domain-containing protein 8	0.00	2.48	1	1	1	1	2.056E7	564	62.3	4.67
Q9FHY8	At5g41950	4.33	4.78	1	2	2	4	6.865E7	565	62.4	4.75
Q9M2Q9	Putative uncharacterized protein T10K17.100	5.61	8.42	3	3	3	3	8.755E7	570	62.4	5.74
B9DHY1	AT3G14067 protein (Fragment)	24.37	6.66	3	3	3	8	2.464E8	601	62.5	6.64
Q9C5C2	Myrosinase 2	14.90	10.05	1	2	3	5	4.249E7	547	62.7	7.44
Q93ZJ6	At2g32240/F22D22.1	2.66	2.29	3	1	1	1	5.967E6	568	62.9	4.98
Q8LE50	Putative splicing factor	5.49	3.39	4	1	1	2	1.656E7	560	63.0	6.84
Q9LD90	H/ACA ribonucleoprotein complex subunit 4	0.00	3.36	1	1	1	1	5.379E7	565	63.0	9.14
Q84K25	Conserved oligomeric Golgi complex component-related protein	5.50	2.46	3	1	1	2	5.963E6	569	63.0	5.17
O49299	Probable phosphoglucomutase, cytoplasmic 1	43.90	15.95	4	2	5	12	7.130E7	583	63.1	6.30
B9DHX2	Phenylalanine ammonia-lyase (Fragment)	8.89	7.43	4	2	3	4	3.399E7	579	63.2	6.43
C0Z361	Chaperonin 60 subunit beta 3, chloroplastic	15.14	13.74	4	6	6	7	1.307E8	597	63.3	5.87
O49485	D-3-phosphoglycerate dehydrogenase 1, chloroplastic	4.64	1.99	1	1	1	3	1.925E8	603	63.3	6.58
Q9SGC1	Probable phosphoglucomutase, cytoplasmic 2	15.30	12.48	4	1	4	7	6.894E7	585	63.4	5.82
Q9FLW9	Plastidial pyruvate kinase 2	8.15	9.15	1	3	3	4	7.944E6	579	63.5	7.03
B9DH71	Ketol-acid reductoisomerase	3.17	1.69	2	1	1	1	4.494E7	591	63.7	6.99
Q9SW96	Asparagine--tRNA ligase, cytoplasmic 1	2.45	3.15	1	1	1	1	2.877E7	572	63.7	5.58
F4K4Y6	RNA recognition motif-containing protein	6.56	3.59	2	1	1	2	1.523E7	585	64.0	5.43
Q0WME1	Putative uncharacterized protein At4g23440 (Fragment)	2.21	2.45	2	1	1	1	5.859E7	571	64.1	9.57
Q8VX13	Protein disulfide isomerase-like 1-3	0.00	2.94	1	1	1	1	7.362E6	579	64.2	4.83
F4HZU9	Calcium-binding EF-hand family protein	0.00	2.42	3	1	1	1	3.723E6	578	64.3	6.79
Q9LYG3	NADP-dependent malic enzyme 2	8.77	7.14	2	2	3	3	2.432E7	588	64.4	6.42
Q5XEP2	Hsp70-Hsp90 organizing protein 2	4.59	3.68	1	2	2	2	2.823E7	571	64.5	6.18
O82762	F17H15.1/F17H15.1	19.47	12.03	3	5	5	9	5.807E7	632	64.6	5.49
Q9XGZ0	NADP-dependent malic enzyme 3	14.65	5.44	2	1	2	6	3.841E7	588	64.6	6.99
Q66GI4	Proteinaceous RNase P 1, chloroplastic/mitochondrial	2.11	2.10	1	1	1	1	4.073E8	572	64.8	9.03
Q9LIR4	Dihydroxy-acid dehydratase, chloroplastic	22.30	6.09	1	2	2	5	7.468E7	608	64.9	6.23
Q9LV77	Asparagine synthetase [glutamine-hydrolyzing] 2	4.70	1.90	1	1	1	2	2.337E7	578	65.0	6.46
Q9SW48	Probable alkaline/neutral invertase B	2.53	2.10	1	1	1	1	1.014E8	571	65.0	6.54
Q9LIK0	Plastidial pyruvate kinase 1, chloroplastic	1.85	1.17	1	1	1	1	2.515E7	596	65.1	5.92
O80763	Probable nucleoredoxin 1	9.55	4.50	1	2	2	4	5.732E7	578	65.1	5.01
Q9SU40	Monocopper oxidase-like protein SKU5	4.61	2.73	1	1	1	2	3.763E7	587	65.6	9.11
Q9FLP0	65-kDa microtubule-associated protein 1	13.61	3.41	2	1	1	4	1.158E8	587	65.7	5.06
Q9SQT8	Bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase, chloroplastic	5.72	4.98	1	2	2	2	5.770E7	603	65.8	6.87
Q9FGX1	ATP-citrate synthase beta chain protein 2	37.24	18.91	3	8	8	13	1.375E8	608	65.8	7.64
Q9C614	Probable L-gulonolactone oxidase 1	0.00	2.52	1	1	1	2	1.682E7	595	65.9	7.06
F4JEJ0	Beta-fructofuranosidase, insoluble isoenzyme CWINV1	5.25	3.27	3	2	2	2	3.052E7	581	65.9	9.13
F4IDD6	Probable phenylalanine--tRNA ligase beta subunit	0.00	1.37	2	1	1	1	1.206E7	584	65.9	5.49

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9SE60	Methylenetetrahydrofolate reductase 1	38.88	21.96	3	5	11	17	2.781E8	592	66.2	5.82
Q9FF55	Protein disulfide isomerase-like 1-4	2.74	7.37	3	3	3	3	2.355E7	597	66.3	4.61
O23246	Arginyl-tRNA synthetase	0.00	1.87	4	2	2	2	2.323E7	589	66.3	6.35
O04130	D-3-phosphoglycerate dehydrogenase 2, chloroplastic	10.25	5.93	1	2	2	6	8.758E7	624	66.4	6.13
F4J7S8	Phosphatidylinositol/phosphatidylcholine transfer protein SFH9	0.00	2.94	3	2	2	2	3.257E7	579	66.6	8.54
C0Z2L4	AT5G52310 protein	12.50	4.16	4	1	1	3	3.199E7	601	66.6	4.51
A8MS68	Dihydrolipoyl dehydrogenase 1, chloroplastic	11.14	16.85	1	3	7	8	1.128E8	623	66.6	8.41
B3H581	DegP2 protease	0.00	2.31	2	1	1	1	1.015E7	606	66.6	5.95
P56820	Eukaryotic translation initiation factor 3 subunit D	1.69	3.05	3	2	2	2	1.360E7	591	66.7	5.77
Q9SJF1	GTPase LSG1-2	2.26	3.57	1	1	1	1	1.251E8	589	66.7	5.85
O80585	Methylenetetrahydrofolate reductase 2	23.51	16.50	1	3	9	17	1.877E8	594	66.8	5.50
A0A090MHY5	Glucose-6-phosphate isomerase	2.65	1.96	2	1	1	1	1.420E7	611	66.8	5.60
F4K1E8	Uncharacterized protein	0.00	3.05	1	1	1	1	1.220E7	590	67.0	8.66
F4K007	Luminal-binding protein 2	47.05	20.07	2	3	9	20	1.454E8	613	67.4	5.29
V9H1C6	Reverse transcriptase (Fragment)	2.51	4.09	1	1	1	1		587	67.4	6.01
Q9C9K3	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha 2	2.40	5.02	1	3	3	4	7.668E7	617	67.5	7.23
Q9LJ94	Genomic DNA, chromosome 3, BAC clone:F16J14	2.28	3.50	1	1	1	1	7.216E6	600	67.5	5.40
Q0WVV6	Calmodulin-binding protein 60 D	1.77	2.16	6	2	2	2	1.529E7	601	67.5	5.48
Q9SPK5	Formate--tetrahydrofolate ligase	6.54	3.47	2	3	3	5	3.275E7	634	67.8	6.71
Q0WWF9	Elongation factor Ts (Fragment)	15.32	10.39	6	2	2	4	4.398E7	616	67.8	4.72
Q8W112	Beta-D-glucan exohydrolase-like protein	9.49	5.45	22	3	3	8	5.244E7	624	67.9	9.13
Q9SCY0	Phosphoglucomutase, chloroplastic	5.79	4.33	1	2	2	3	9.614E6	623	67.9	5.74
Q06850	Calcium-dependent protein kinase 1	3.57	3.28	1	1	1	2	3.449E7	610	68.2	5.50
Q9FFD4	Protein kinase-like protein	0.00	2.14	2	1	1	1	1.610E7	608	68.5	4.73
F4JUM3	Sec23/sec24-like transport protein	5.07	3.06	5	1	1	1	5.473E7	620	68.6	6.30
P42731	Polyadenylate-binding protein 2	4.92	5.72	1	2	2	2	3.595E7	629	68.6	8.21
O23654	V-type proton ATPase catalytic subunit A	10.04	6.10	3	4	4	5	1.127E8	623	68.8	5.24
Q9SAV1	Myrosinase-binding protein 2	4.86	4.52	5	2	2	3	3.596E7	642	68.8	6.79
O23144	Proton pump-interactor 1	1.83	1.63	1	1	1	1	1.099E7	612	68.8	9.03
Q9M0G0	Gamma-glutamyltranspeptidase 3	7.62	4.08	1	2	2	3	8.721E7	637	69.1	6.21
Q9FY49	Leukotriene A-4 hydrolase homolog	1.90	4.87	1	2	2	2	1.619E7	616	69.2	5.24
Q9SIU0	NAD-dependent malic enzyme 1, mitochondrial	6.95	1.93	1	1	1	3	4.050E7	623	69.6	5.45
O82663	Succinate dehydrogenase [ubiquinone] flavoprotein subunit 1, mitochondrial	29.00	8.52	2	3	3	8	6.865E7	634	69.6	6.29
F4IRA9	Protein IQ-domain 29	2.38	3.93	2	1	1	1	1.191E7	636	69.7	9.60
Q56ZQ3	Vacuolar-sorting receptor 4	2.74	2.55	2	1	1	1	1.076E7	628	69.8	5.48
Q94JM0	AT5g66420/K1F13_7	2.09	1.37	2	1	1	1	1.674E7	655	70.0	6.06
Q9LDL2	Putative uncharacterized protein AT4g26600	0.00	0.96	2	1	1	1	3.411E7	626	70.6	5.67
Q9LHA8	Probable mediator of RNA polymerase II transcription subunit 37c	16.32	8.00	2	1	4	6	3.505E7	650	71.1	5.25
Q9LP77	Probable inactive receptor kinase At1g48480	3.06	2.75	1	1	1	1	2.088E6	655	71.1	8.10
O65719	Heat shock 70 kDa protein 3	61.18	12.94	1	2	6	18	5.020E8	649	71.1	5.07
Q9CA83	NADP-dependent malic enzyme 4, chloroplastic	15.82	7.43	3	2	3	6	2.818E7	646	71.1	6.58
Q9XI20	Alpha/beta-hydrolase domain-containing protein	0.00	1.08	1	1	1	1	1.325E7	648	71.2	9.20
P22954	Probable mediator of RNA polymerase II transcription subunit 37c	32.35	14.85	2	3	7	12	7.188E7	653	71.3	5.12
O22173	Polyadenylate-binding protein 4	5.52	2.27	1	2	2	2	2.052E7	662	71.6	6.86

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
O82392	Phosphomethylpyrimidine synthase, chloroplastic	19.45	4.50	1	1	1	4	1.071E8	644	71.9	6.44
Q9FNF2	Starch synthase 1, chloroplastic/amyloplastic	2.56	2.30	1	1	1	1	3.149E7	652	72.1	6.46
Q9SFF5	F10A16.20 protein	14.51	5.61	3	2	2	4	3.028E7	660	72.1	4.34
B9DFC8	AT5G55660 protein (Fragment)	3.36	1.56	3	1	1	2	1.369E7	643	72.5	5.00
Q9SGR9	T23E18.7	6.11	2.00	4	2	2	3	4.901E7	650	72.8	5.49
Q8L860	ENTH/VHS/GAT family protein	3.29	3.41	3	1	1	1	3.539E7	675	72.9	5.19
Q9LDZ0	Heat shock 70 kDa protein 10, mitochondrial	5.60	3.67	1	1	2	3	1.403E8	682	72.9	5.78
Q9FXD4	Signal recognition particle subunit SRP72	4.88	2.26	1	1	1	2	2.100E7	664	73.0	8.82
Q9MA55	Acyl-CoA-binding domain-containing protein 4	9.06	6.89	1	3	3	3	1.120E7	668	73.0	5.30
Q8GUM2	Heat shock 70 kDa protein 9, mitochondrial	2.79	2.49	2	1	2	4	1.981E8	682	73.0	5.62
Q9FM96	Glucosidase 2 subunit beta	9.21	3.71	1	1	1	2	7.862E6	647	73.2	4.79
Q9LKR3	Mediator of RNA polymerase II transcription subunit 37a	45.39	16.14	2	1	7	14	1.454E8	669	73.6	5.17
Q39041	Acid beta-fructofuranosidase 4, vacuolar	8.47	3.01	1	2	2	4	3.320E7	664	73.8	5.63
Q0WM35	Putative uncharacterized protein At3g63460 (Fragment)	8.64	4.12	5	2	2	3	1.141E7	679	73.8	4.93
Q9ZU23	Jacalin-related lectin 5	35.11	13.29	1	4	4	8	1.229E7	730	73.9	5.16
Q9ZQP2	Putative peroxisomal acyl-coenzyme A oxidase 1.2	1.77	1.51	1	1	1	1	5.367E7	664	74.3	8.73
O65202	Peroxisomal acyl-coenzyme A oxidase 1	10.17	7.68	2	4	4	7	1.370E8	664	74.3	7.66
Q9T0A0	Long chain acyl-CoA synthetase 4	0.00	2.10	1	1	1	1	4.672E7	666	74.5	5.87
Q93YP3	Subtilisin proteinase-like	2.21	1.85	3	2	2	2	3.657E7	703	74.5	8.29
Q9SG80	Alpha-L-arabinofuranosidase 1	10.21	2.80	1	1	1	2	1.456E7	678	75.0	5.77
Q93Z12	AT4g33010/F26P21_130	16.34	8.50	3	2	4	8	5.900E7	694	75.0	6.52
Q9SSG3	HIPL1 protein	1.70	3.88	1	2	2	2	3.573E7	695	75.2	5.52
Q9LKS4	Putative uncharacterized protein T15F17.d	0.00	3.25	1	1	1	2	2.751E7	677	76.0	6.62
Q93XY1	At5g53620	1.73	1.17	2	1	1	1	1.639E7	682	76.3	5.83
Q9STW6	Heat shock 70 kDa protein 6, chloroplastic	9.11	4.87	2	3	3	4	6.076E7	718	76.5	5.20
Q9FKI0	Fimbrin-5	2.00	1.75	1	1	1	1	5.745E6	687	76.7	5.27
Q9SS45	Phenylalanine ammonia-lyase 4	9.22	5.52	3	1	2	3	3.878E7	707	76.9	6.27
P0CB21	Uncharacterized protein At4g26450	9.80	3.11	1	1	1	3	1.421E7	708	77.5	5.39
Q9ZU46	Leucine-rich repeat receptor-like protein kinase	0.00	1.12	1	1	1	1	7.131E8	716	78.3	6.11
P35510	Phenylalanine ammonia-lyase 1	27.80	13.79	7	5	6	10	1.269E8	725	78.7	6.30
Q949P2	Probable cytosolic oligopeptidase A	9.35	4.99	1	1	3	4	8.307E7	701	79.0	5.63
Q9MAP4	Fourth of four adjacent putative subtilase family	0.00	3.13	1	1	1	1	2.964E7	734	79.1	6.52
O65351	Subtilisin-like protease SBT1.7	55.28	15.98	2	6	6	16	1.380E8	757	79.4	6.35
F4JBY2	Transketolase	31.09	15.68	2	8	8	13	3.440E8	740	79.8	6.43
F4IW47	Transketolase-2, chloroplastic	0.00	2.43	1	1	1	1	2.640E7	741	79.9	6.58
P51818	Heat shock protein 90-3	16.74	8.44	8	6	6	10	3.838E8	699	80.0	5.05
Q8H1H9	Primary amine oxidase	2.22	1.26	1	1	1	1	3.776E6	712	80.1	6.43
Q9M1H3	ABC transporter F family member 4	3.41	1.80	1	1	1	1	2.223E7	723	80.4	6.32
O81908	Pentatricopeptide repeat-containing protein At1g02060, chloroplastic	0.00	1.83	1	1	1	1	1.073E7	710	80.8	8.40
O04630	Threonine--tRNA ligase, mitochondrial	4.34	2.68	1	2	2	2	2.277E7	709	80.9	6.98
O49607	Subtilisin-like protease SBT1.6	12.30	3.80	1	3	3	5	2.866E7	764	81.0	9.35
Q9ZVD0	Serrate RNA effector molecule	0.00	1.53	1	1	1	1	7.848E6	720	81.0	8.38
Q9LNU1	CO(2)-response secreted protease	2.78	1.56	1	1	1	2	9.065E7	769	81.4	6.00
O64760	At2g34970	1.80	2.05	1	1	1	1		730	81.8	4.68
Q8LPT3	Membrane protein of ER body-like protein	0.00	3.15	1	1	1	1	1.532E6	761	82.5	4.79
Q9LXD6	Beta-D-xylosidase 3	17.13	7.24	3	6	7	9	1.393E8	773	83.2	7.46

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q8GXD0	Putative acetyl-CoA synthetase	0.00	2.22	2	1	1	1	2.540E8	767	83.3	7.81
F4JXC5	Subtilisin-like protease SBT5.4	6.71	3.34	1	3	3	4	6.089E7	778	83.4	8.72
Q9SGZ5	Probable beta-D-xylosidase 7	5.82	3.78	1	2	2	2	2.971E7	767	83.8	8.03
Q9LUG1	Protein transport protein Sec23	0.00	1.70	2	1	1	1	3.741E7	765	84.0	5.48
F4J9W8	Transducin family protein / WD-40 repeat family protein	0.00	1.97	2	1	1	1	2.701E7	760	84.2	5.58
O50008	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase 1	47.47	8.24	4	2	7	16	5.827E8	765	84.3	6.51
F4ICX9	TSK-associating protein 1	2.44	2.50	1	1	1	1	6.515E6	759	84.3	4.65
Q9SRV5	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase 2	39.94	7.58	7	1	6	14	5.784E8	765	84.5	6.51
F4HWL4	Probably tyrosine-tRNA synthetase	0.00	5.48	2	3	3	4	1.480E7	748	84.7	6.61
Q9SAB1	Heat shock 70 kDa protein 16	3.99	3.28	1	2	2	2	1.135E7	763	85.2	5.71
Q9LD43	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha, chloroplastic	15.04	7.80	1	4	4	7	7.155E7	769	85.3	5.82
Q9LJN4	Probable beta-D-xylosidase 5	8.30	6.02	1	3	4	7	9.984E7	781	85.4	8.51
Q9M1Z2	LETM1-like protein	0.00	1.72	2	1	1	1	7.855E6	755	85.7	6.71
A0A097PNI8	ABH1 (Fragment)	3.10	2.13	2	1	1	1	3.237E7	750	86.0	5.55
F4II29	Cytoplasmic tRNA export protein	0.00	2.76	2	1	1	1	3.677E7	798	86.0	7.85
Q9SI75	Elongation factor G, chloroplastic	11.95	4.09	3	3	3	5	6.529E7	783	86.0	5.60
Q94AH6	Cullin-1	0.00	1.49	1	1	1	3	1.036E7	738	86.2	7.14
Q8H0S3	At1g05520/T25N20_16	1.82	1.66	3	1	1	1	3.425E6	783	86.3	5.64
Q8L831	Nudix hydrolase 3	2.45	1.55	1	1	1	1	8.128E6	772	86.8	5.44
Q9FXI5	Protein IQ-DOMAIN 32	13.33	2.52	1	1	1	5	2.131E7	794	86.8	5.30
Q9STW7	Putative uncharacterized protein AT4g24270	3.62	1.95	4	1	1	1	1.746E7	768	87.1	5.76
O23638	Heat shock protein	11.32	4.30	6	2	2	6	3.300E7	768	87.4	5.05
Q8RXF1	Probable splicing factor 3A subunit 1	0.00	2.93	1	1	1	1	1.252E7	785	87.5	5.15
Q42572	DNA ligase 1	2.54	1.39	1	1	1	1	5.353E6	790	87.7	7.99
Q8L6Y1	Ubiquitin carboxyl-terminal hydrolase 14	0.00	1.51	1	1	1	1	1.800E7	797	88.3	5.21
Q94AM1	Organellar oligopeptidase A, chloroplastic/mitochondrial	12.03	4.42	1	1	3	5	8.624E7	791	88.7	6.28
P54609	Cell division control protein 48 homolog A	56.52	18.91	3	11	11	24	1.701E8	809	89.3	5.25
Q8L601	Putative uncharacterized protein At2g22400	2.70	2.10	2	1	1	1	4.159E6	808	89.8	6.06
Q9SVN5	Probable methionine--tRNA ligase	4.98	2.13	1	1	1	2	9.716E6	797	89.8	7.01
B9DI51	AT1G29900 protein (Fragment)	4.50	1.33	2	1	1	3	1.509E7	825	90.5	5.85
Q9LVI0	Genomic DNA, chromosome 3, P1 clone: MEB5	4.01	2.04	2	1	1	1	1.084E7	833	91.0	5.78
F4HQD4	Heat shock 70 kDa protein 15	6.26	6.02	2	2	5	5	1.300E8	831	91.6	5.19
Q9S7C0	Heat shock 70 kDa protein 14	6.26	5.29	1	1	4	4	1.250E8	831	91.7	5.25
Q9FMQ1	Pentatricopeptide repeat-containing protein At5g12100, mitochondrial	0.00	1.96	1	1	1	1	2.046E7	816	91.9	6.28
Q8L745	AT5g42220/K5J14_2	4.47	1.25	2	1	1	2	4.201E7	879	92.5	5.22
F4IIQ3	Beta-galactosidase	8.60	3.31	3	3	3	5	4.576E7	846	92.5	7.85
P49040	Sucrose synthase 1	10.02	6.06	1	5	5	7	2.204E7	808	92.9	6.20
Q9SCP3	Nodulin / glutamate-ammonia ligase-like protein	2.02	1.42	3	1	1	1	6.401E6	845	93.5	5.90
Q9ASR1	At1g56070/T6H22_13	32.71	17.08	10	16	16	22	4.216E8	843	93.8	6.25
F4JQ55	HSP90-like protein GRP94	27.62	7.17	3	5	5	9	2.287E8	823	94.1	5.01
Q9LXU3	Putative uncharacterized protein T24H18_130	2.78	2.33	3	1	1	3	1.217E7	860	95.8	7.17
Q9LFE4	WEB family protein At5g16730, chloroplastic	5.43	3.05	1	1	1	1		853	96.1	4.94
F4JMJ1	Heat shock 70 kDa protein 17	10.89	4.84	2	4	4	6	1.257E8	867	96.7	6.14

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q6XJG8	26S proteasome non-ATPase regulatory subunit 2 homolog B	15.56	4.94	2	3	3	5	1.385E8	891	97.9	5.21
Q06327	Linoleate 9S-lipoxygenase 1	104.98	19.32	1	15	15	39	4.694E8	859	98.0	5.52
Q42560	Aconitate hydratase 1	31.50	11.58	2	8	11	16	2.701E8	898	98.1	6.40
Q8VZH2	Aminopeptidase M1	4.13	3.30	1	2	2	3	7.839E6	879	98.1	5.50
Q0WW26	Coatomer subunit gamma	0.00	1.69	1	1	1	1	5.067E6	886	98.4	5.16
Q9SA18	Bifunctional aspartokinase/homoserine dehydrogenase 1, chloroplastic	6.64	2.31	1	1	1	2	6.593E6	911	99.3	6.77
P93831	Histone-lysine N-methyltransferase CLF	0.00	2.11	1	1	1	1	1.438E6	902	100.3	8.60
P38418	Lipoxygenase 2, chloroplastic	13.09	7.25	1	4	4	6	1.289E8	896	102.0	5.62
Q9S7Y7	Alpha-xylosidase 1	7.84	1.86	1	1	1	2	6.686E6	915	102.3	6.77
Q9ZVD5	Protein argonaute 4	5.17	1.52	1	2	2	2	1.506E7	924	102.8	8.88
P42762	Chaperone protein ClpD, chloroplastic	0.00	1.59	1	1	1	1	4.168E6	945	103.2	6.23
Q9FI56	Chaperone protein ClpC1, chloroplastic	12.95	6.24	4	5	5	11	1.837E8	929	103.4	6.77
Q9FT90	Putative uncharacterized protein F8L15_180	6.50	3.38	2	2	2	3	2.719E7	917	103.9	7.58
O22864	Protein NLP8	0.00	1.06	1	1	1	1	1.261E7	947	104.8	5.67
O23404	Pyruvate, phosphate dikinase 1, chloroplastic	0.00	1.77	1	1	1	1	3.529E7	963	105.1	6.38
F4I4Z2	Alanine-tRNA ligase	2.19	2.80	5	3	3	3	3.019E7	963	105.9	5.88
Q8VZF3	Probable glutamyl endopeptidase, chloroplastic	8.30	4.06	7	2	3	5	4.596E7	960	106.0	6.42
O81645	Villin-3	9.15	3.83	2	2	2	4	1.642E7	965	106.3	5.85
B3H6B8	Alpha-mannosidase	0.00	2.33	3	1	1	2	2.902E7	943	106.4	7.11
F4IFQ0	Serine/threonine-protein phosphatase	25.04	6.59	2	3	3	6	3.104E7	1001	107.0	5.81
O81644	Villin-2	29.86	8.91	1	6	6	11	1.088E8	976	107.8	5.29
F4HW26	Serine/threonine-protein phosphatase	7.59	1.58	4	1	1	2	4.217E6	1013	108.0	5.80
Q9SIB9	Aconitate hydratase 2, mitochondrial	31.50	12.83	2	9	12	20	2.736E8	990	108.1	7.17
Q0WVL7	Golgin candidate 5	8.66	1.57	2	1	1	2	1.198E7	956	108.3	4.83
Q94A28	Aconitate hydratase 3, mitochondrial	0.00	0.80	1	1	1	1	7.654E6	995	108.4	7.15
Q9LIB2	Alpha-glucan phosphorylase 1	10.26	2.39	1	1	1	3	3.694E7	962	108.5	5.47
F4JWP9	Ribosomal protein S5/Elongation factor G/III/V family protein	0.00	0.51	1	1	1	1	1.753E7	973	108.7	5.24
Q9MAT0	26S proteasome non-ATPase regulatory subunit 1 homolog B	7.73	5.29	2	3	3	5	2.415E7	1001	108.8	5.35
Q9LF37	Chaperone protein ClpB3, chloroplastic	11.84	5.99	4	6	6	6	6.889E7	968	108.9	6.23
O65570	Villin-4	12.22	3.49	1	3	3	5	3.815E7	974	109.3	6.13
Q9SHI1	Translation initiation factor IF-2, chloroplastic	0.00	1.36	1	1	1	1	1.272E7	1026	109.7	7.44
Q5GM68	Phosphoenolpyruvate carboxylase 2	1.72	2.39	4	2	2	2	3.099E7	963	109.7	5.80
Q8RXD9	4-alpha-glucanotransferase DPE2	5.66	3.46	1	2	2	2	4.690E7	955	109.7	5.82
O22941	Zinc-metallopeptidase, peroxisomal	2.89	1.44	1	1	1	1	5.423E6	970	110.9	6.04
Q9XEA0	AT4g04350/T19B17_7	0.00	1.34	1	1	1	1	2.139E7	973	111.0	6.30
O80988	Glycine dehydrogenase (decarboxylating) 2, mitochondrial	7.85	3.45	1	1	3	4	4.160E7	1044	113.7	6.65
Q9SFU0	Protein transport protein Sec24-like At3g07100	0.00	0.67	1	1	1	1	2.079E7	1038	113.9	7.91
Q9LD55	Eukaryotic translation initiation factor 3 subunit A	2.08	1.32	1	1	1	2	6.197E7	987	114.2	9.17
F4J1Y2	Transport protein sec24-like CEF	4.59	1.03	2	1	1	2	3.541E7	1069	114.6	7.06
F4K3X1	Ubiquitin carboxyl-terminal hydrolase	0.00	1.02	4	1	1	1	3.595E7	985	115.1	5.68
O04379	Protein argonaute 1	2.76	1.24	1	1	1	1	4.377E7	1048	116.1	9.29
Q94BT0	Sucrose-phosphate synthase 1	1.86	0.58	2	1	1	1	3.618E7	1043	117.2	6.43
Q9LF41	Probable ubiquitin conjugation factor E4	4.42	1.16	1	1	1	2	5.507E7	1038	117.5	5.62
F4JA10	Zinc metalloprotease pitrilysin subfamily A	4.13	4.68	3	5	5	5	4.780E7	1069	119.7	5.63
P93028	Ubiquitin-activating enzyme E1 1	8.36	4.07	2	4	4	6	7.560E7	1080	120.2	5.27
Q6NPT2	At1g14610	0.00	0.85	2	1	1	1	1.868E7	1064	120.9	6.60

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
F4J043	WUS-interacting protein 2	5.40	4.09	5	1	2	2	3.490E7	1125	122.8	7.40
Q93VS8	Protein EMBRYO DEFECTIVE 2734	2.60	2.24	1	1	1	2	2.192E7	1116	123.7	4.87
Q0WV90	Topless-related protein 1	8.58	3.66	4	1	2	2	2.499E7	1120	124.0	7.14
Q94AI7	Protein TOPLESS	16.74	4.95	4	2	3	4	2.499E7	1131	124.2	7.24
F4HW65	Protein STICHEL-like 1	1.81	0.72	1	1	1	1	2.099E7	1116	124.8	8.53
Q9S7I6	LRR receptor-like serine/threonine-protein kinase RPK2	1.88	0.78	1	1	1	1	1.918E7	1151	125.2	6.90
O81449	Putative transposon protein	0.00	0.63	3	1	1	1	3.037E7	1104	126.3	8.02
Q9SUG7	Putative uncharacterized protein AT4g30790	0.00	0.87	1	1	1	1	4.524E6	1148	129.1	5.74
Q9LPP9	F15H18.21	0.00	1.00	5	1	1	1	2.583E7	1197	129.4	5.91
Q9XIH2	Putative uncharacterized protein At2g16180	0.00	0.49	1	1	1	1	1.323E7	1218	136.3	5.35
Q94A40	Coatomer subunit alpha-1	8.11	3.95	3	3	3	4	1.260E7	1216	136.5	6.98
Q9FJR0	Regulator of nonsense transcripts 1 homolog	7.53	1.99	1	1	1	2	2.397E7	1254	136.8	6.48
Q9FIZ7	5-oxoprolinase	2.44	3.08	1	2	2	3	2.883E7	1266	137.4	5.86
Q9SRD2	Putative translation initiation factor IF-2; 73082-68138	4.33	2.42	5	3	3	5	3.887E7	1280	140.6	5.40
Q9LTT8	Enhancer of mRNA-decapping protein 4	11.48	2.38	2	2	2	3	1.715E7	1344	145.6	6.10
Q9SUK4	Disease resistance RPP5 like protein	0.00	0.84	3	1	1	1	1.485E7	1304	147.6	5.55
Q9FKN5	Gb|AAC35233.1|	4.67	1.34	1	1	1	1	1.268E7	1342	149.5	4.31
F4HWC3	Uncharacterized protein	0.00	2.42	2	2	2	3	7.104E7	1364	150.1	4.75
F4JVN6	Tripeptidyl-peptidase 2	4.07	1.30	1	2	2	2	2.557E7	1380	152.3	6.07
O82493	T12H20.12 protein	0.00	0.86	2	1	1	1	8.565E6	1392	157.1	8.24
Q9FGR0	Cleavage and polyadenylation specificity factor subunit 1	2.66	0.97	1	1	1	1	2.975E6	1442	158.0	6.14
O81283	Translocase of chloroplast 159, chloroplastic	13.01	3.46	1	3	3	5	2.677E7	1503	160.7	4.55
Q9ZNZ7	Ferredoxin-dependent glutamate synthase 1, chloroplastic/mitochondrial	6.55	1.79	3	1	3	3	2.796E7	1622	176.6	6.32
Q9T0P4	Ferredoxin-dependent glutamate synthase 2, chloroplastic	9.91	3.01	2	2	4	6	9.897E7	1629	177.6	7.01
Q0WL80	UDP-glucose:glycoprotein glucosyltransferase	2.27	0.93	2	2	2	2	9.275E6	1613	181.7	5.83
O23576	Putative uncharacterized protein AT4g17330	2.46	0.87	2	1	1	1	7.985E6	1732	184.7	4.87
A8MR97	Eukaryotic translation initiation factor 4G	16.87	4.47	3	4	4	5	1.625E7	1723	187.4	7.71
O80815	T8F5.22 protein	6.79	1.54	1	1	1	2	2.666E7	1684	189.3	5.24
F4IXW2	Brefeldin A-inhibited guanine nucleotide-exchange protein 5	20.10	2.19	1	2	2	4	2.565E7	1739	192.8	5.58
Q0WLB5	Clathrin heavy chain 2	10.10	0.76	2	1	1	3	1.497E7	1703	193.1	5.38
O49470	Resistance protein RPP5-like	2.12	0.87	2	1	1	1	3.763E7	1715	193.4	8.31
Q9LPC5	Brefeldin A-inhibited guanine nucleotide-exchange protein 3	0.00	1.31	1	1	1	1	2.148E7	1750	194.8	5.64
F4I5Q6	Myosin-7	0.00	0.40	1	1	1	1	6.395E6	1730	196.0	5.57
F4JKH6	Tetratricopeptide repeat domain protein	9.03	1.54	6	2	2	3	2.500E7	1819	199.0	6.16
Q9LV03	Glutamate synthase 1 [NADH], chloroplastic	2.45	0.95	1	2	2	2	6.287E6	2208	241.7	6.37
Q9SSD2	F18B13.15 protein	0.00	0.55	1	1	1	1	4.785E6	2359	275.3	8.79
F4IHS2	Chromatin structure-remodeling complex protein SYD	2.57	0.59	1	1	1	2	8.138E6	3574	389.6	4.87
Q9SRU2	Auxin transport protein BIG	0.00	0.29	1	1	1	1	4.835E7	5098	567.5	5.95

Appendix 3 Supplementary information for Chapter 5

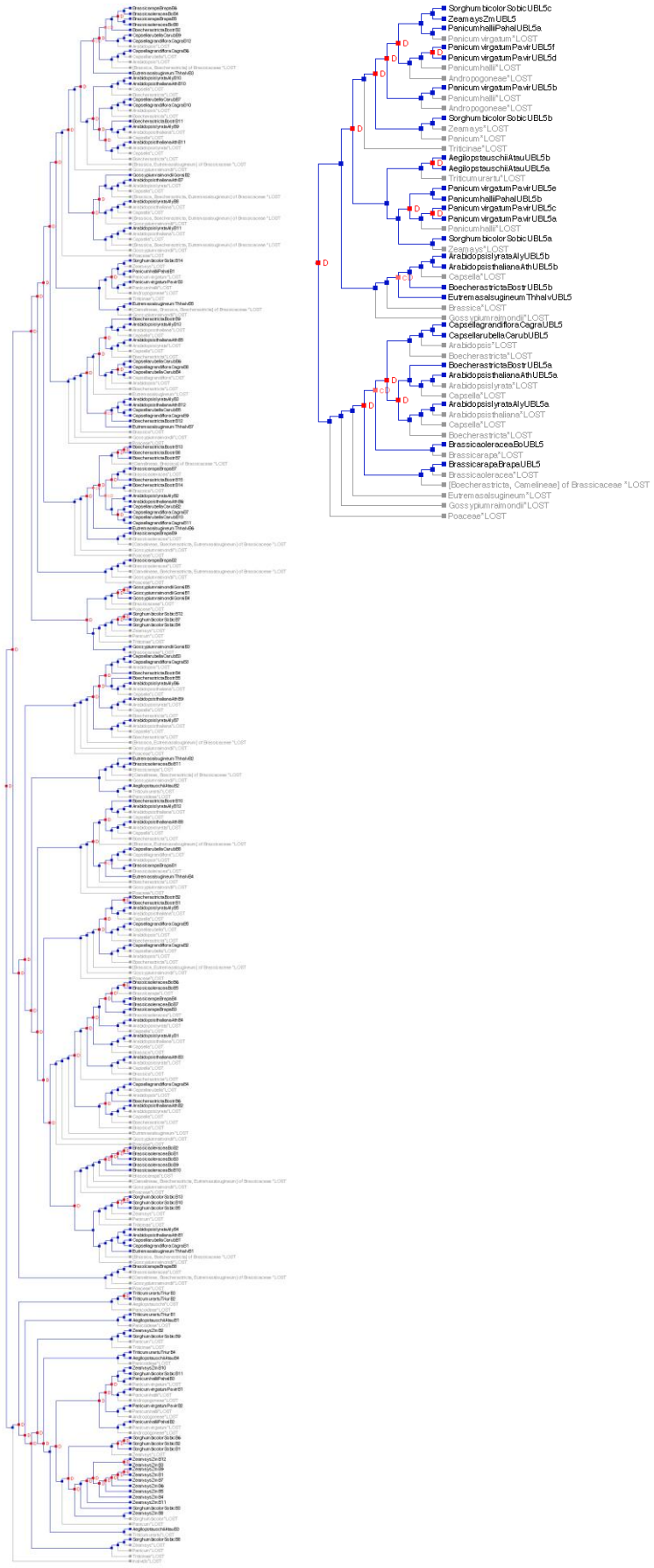


Figure S5.1 | Reconciliation of gene tree and species tree. The two reconciliated trees were built with gene tree of 134 *PCP-B* homologues and their species tree (a) and gene tree of 24 *UBL5* homologues and their species tree (b), respectively. Nodes labeled with 'D' represent possible gene duplication events. The branches shown as grey represent possible gene deletion events.

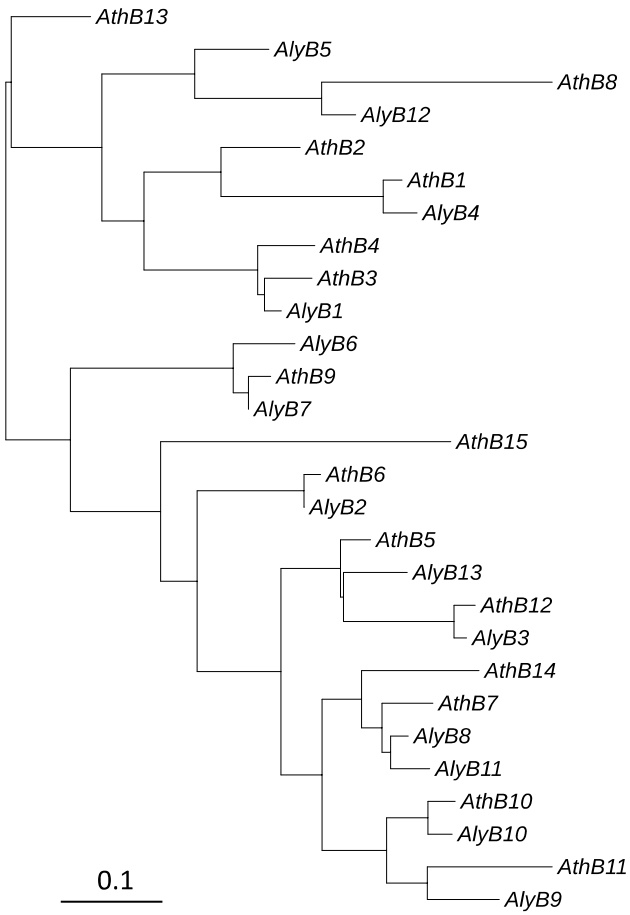
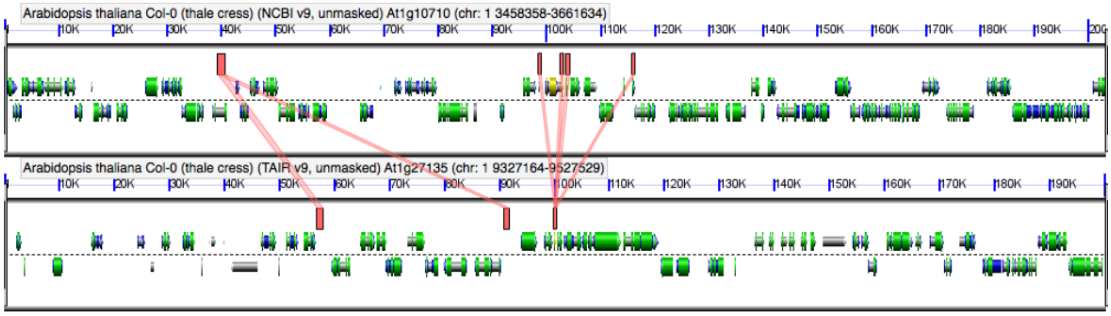
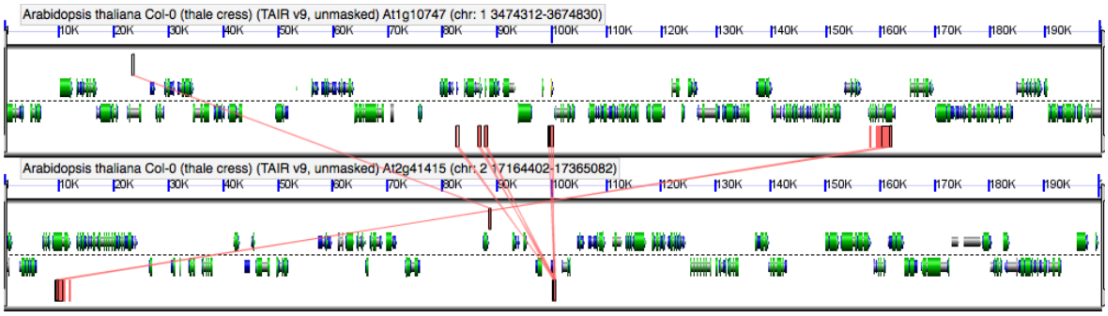


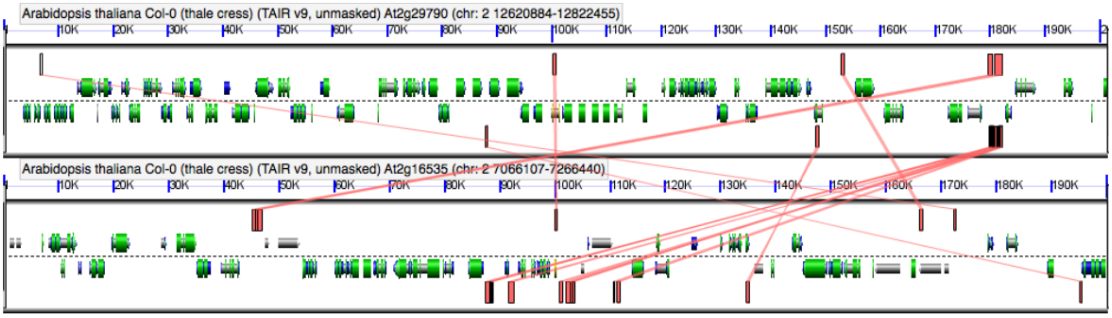
Figure S5.2 | Phylogeny of *PCPBL* genes in *A. thaliana* and *A. lyrata*. Branch length is scaled to the bar defined as 0.1 nucleotide substitutions per codon.



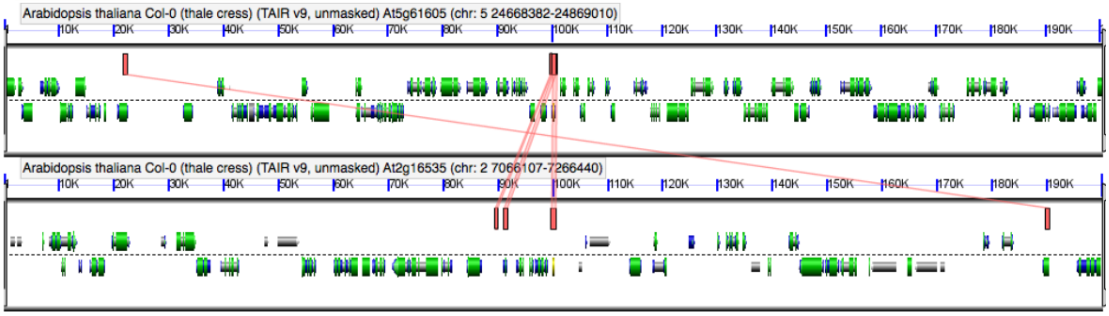
At1g10710 (AthB14): At1g27135 (AthB10)



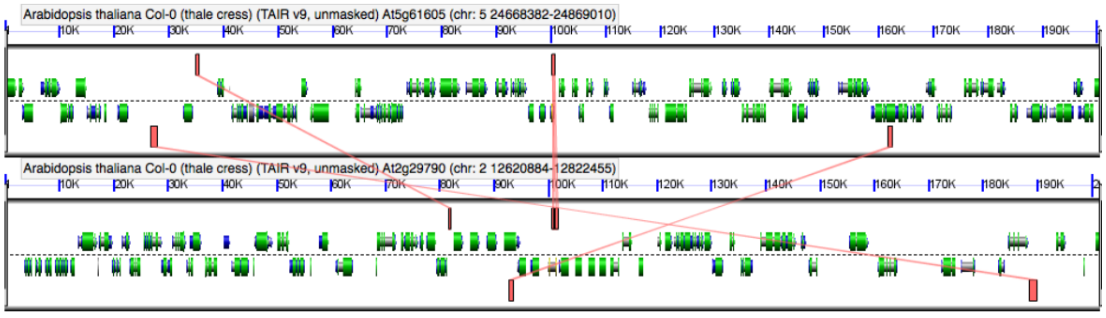
At1g10747 (AthB5) vs. At2g41415 (AthB12)



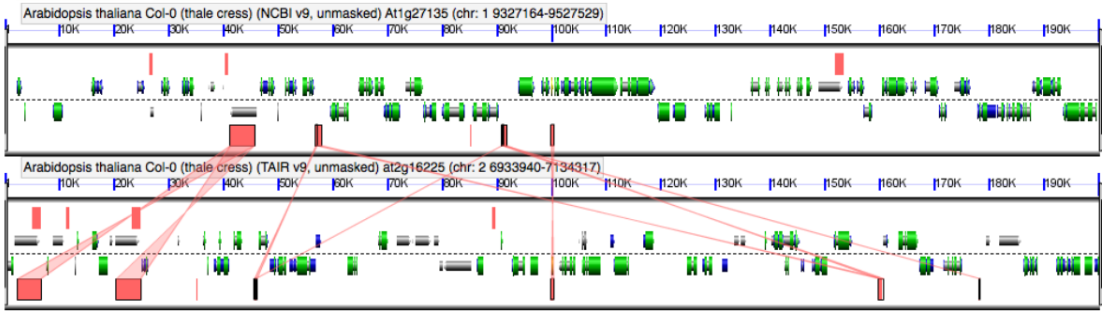
At2g29790 (AthB2) vs. At2g16535 (AthB3)



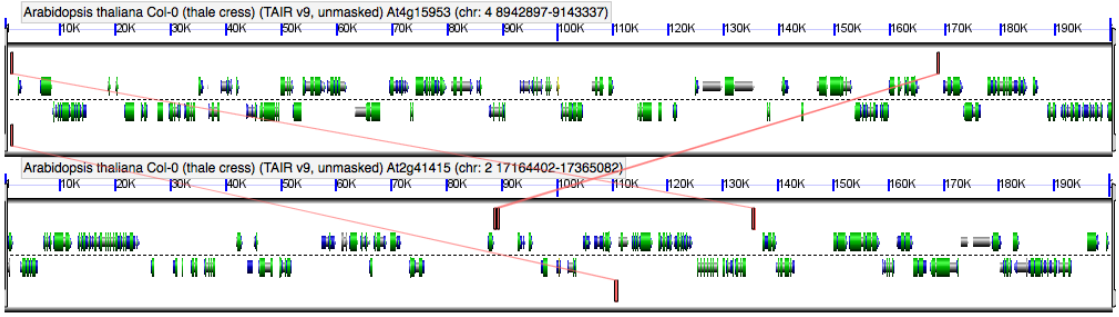
At5g61605 (AthB1) vs. At2g16535 (AthB3)



At5g61605 (AthB1) vs. At2g29790 (AthB2)

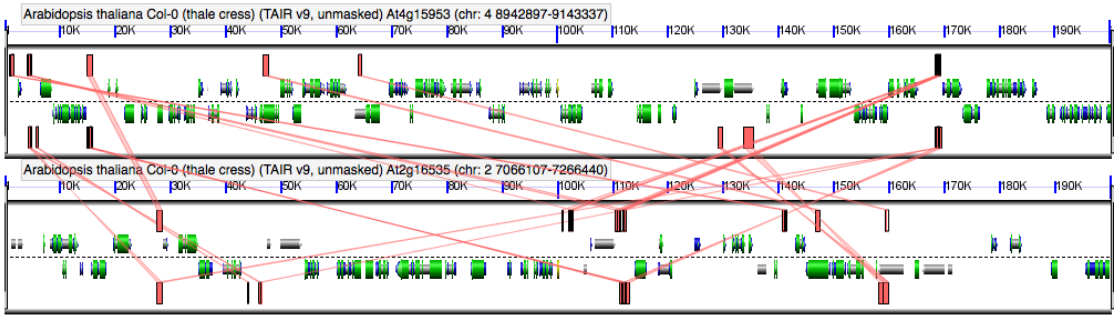


At1g27135 (AthB10) vs. At2g16225 (AthB11)

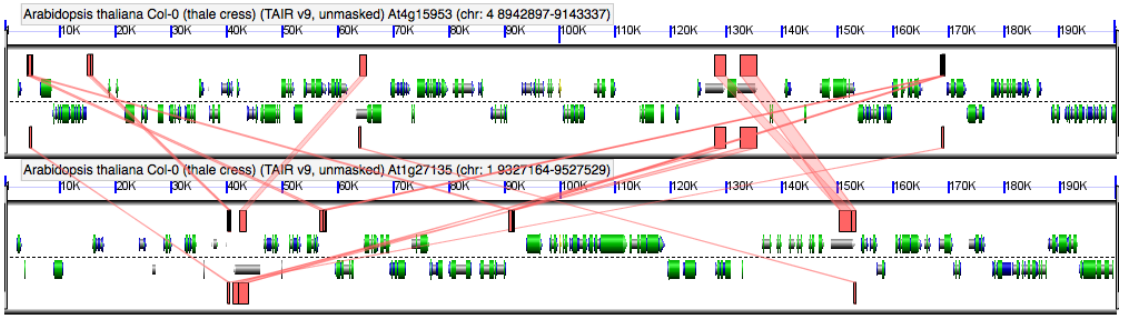


At4g15953 (AthB8): At2g41415 (AthB12)

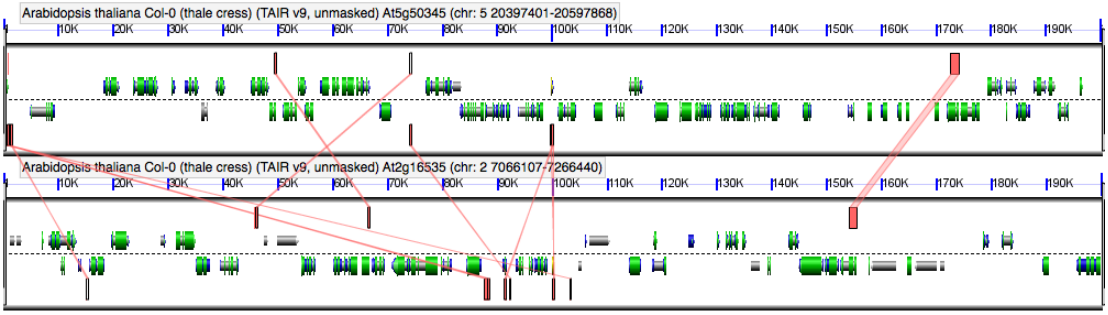
At4g15953 (AthB8): At2g29790 (AthB2)



At4g15953 (AthB8): At2g16535 (AthB3)



At4g15953 (AthB8): At1g27135 (AthB10)



At5g50345 (AthB9): At2g16535 (AthB3)

Figure S5.3 | Comparisons of genomic regions around *PCPBLs* on the genome of *Arabidopsis thaliana.* Reference *PCPBL* genes are marked as yellow (aligned with 100K mark). The high-scoring segment pairs (HSP) identified by GEvo are highlighted by pink bars and connected by wedges. The default score threshold was 3000.

Appendix 4 Supplementary information for Chapter 6

Table S6.1 | Statistically validated* hits from mass spectrometry analysis of proteins obtained from *Arabidopsis thaliana* pollen coat. Score: the sum of the scores of the individual peptides; Coverage: the percentage of the protein sequence covered by identified peptides; Proteins: the number of identified proteins in the protein group of a master protein; Unique peptides: the number of peptide sequences unique to a protein group; Peptides: the number of distinct peptide sequences in the protein group; PSMs: the number of peptide-spectrum match (PSM) for the protein, including those redundantly identified; Area: the average area of the three unique peptides with the largest peak area; pI: theoretically calculated isoelectric point, which is the pH at which a particular molecule carries no net electrical charge. The PSMs were statistically validated to avoid false positives by using the False discovery rate (FDR)-controlling procedure (for details see 2.9.4).

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
B6DVF1	KIN1	126.29	83.33	2	6	8	42	5.180E7	66	6.4	9.47
P31169	Stress-induced protein KIN2	133.14	80.30	1	4	6	43	5.458E8	66	6.5	9.13
Q9FHZ4	ABA-inducible protein-like	73.40	43.28	2	3	3	27	2.483E6	67	7.0	9.47
Q9FKR1	At5g38760	106.53	73.13	3	4	4	36	7.535E7	67	7.0	9.47
Q9M892	AT3g02480/F16B3_11	96.79	48.53	1	4	4	26	4.132E7	68	7.1	9.04
Q9ZUL8	Defensin-like protein 10	32.84	15.07	1	1	1	17	7.313E8	73	7.7	8.51
B2CU06	CSD2 (Fragment)	11.56	47.37	2	1	1	3	1.164E7	76	8.0	4.97
C0Z3J4	AT3G62290 protein	0.00	11.11	6	1	1	1		72	8.1	5.00
Q1G3R6	EMBRYO SURROUNDING FACTOR 1-like protein 4	14.21	32.43	1	1	1	3	1.539E7	74	8.3	7.15
Q9C947	Defensin-like protein 5	11.59	9.21	1	1	1	6	3.391E8	76	8.3	8.13
A8MR88	EMBRYO SURROUNDING FACTOR 1-like protein 8	36.85	42.11	1	4	4	18	5.273E7	76	8.3	7.47
P82739	Defensin-like protein 159	142.58	53.25	1	4	4	57	3.821E8	77	8.4	7.83
O48776	Protein RALF-like 17	4.14	20.78	1	1	1	1	4.047E7	77	8.4	4.41
Q8S8R1	Expressed protein	0.00	26.51	1	1	1	1	1.666E7	83	8.5	4.61
Q2V2W7	Putative defensin-like protein 274	12.51	32.91	1	1	1	4	4.638E7	79	8.5	5.73
Q9LNQ6	At1g17490/F1L3_4	1.89	9.09	1	1	1	1	5.457E6	77	8.5	5.25
A8MQY8	EMBRYO SURROUNDING FACTOR 1-like protein 9	108.95	49.35	1	5	5	47	1.134E9	77	8.6	8.50
Q8S8H3	Defensin-like protein 149	123.42	30.77	1	3	3	53	4.556E7	78	8.6	8.94
Q8S8H9	Defensin-like protein 144	22.07	27.16	1	3	3	10	6.632E6	81	8.8	9.26
Q2V466	Putative defensin-like protein 191	5.37	32.93	1	1	1	4	2.197E6	82	8.9	9.16
P82729	Putative defensin-like protein 137	16.27	20.25	1	1	1	6	3.339E6	79	8.9	9.00
P82748	Putative defensin-like protein 133	7.33	28.21	1	2	2	7	4.811E7	78	9.0	9.00
Q42083	Gene GF14 (Fragment)	8.47	18.99	4	1	1	3	5.881E5	79	9.0	5.01
Q9LF54	Probable calcium-binding protein CML29	2.99	22.89	1	1	1	1	2.328E6	83	9.0	4.70
P82746	Defensin-like protein 153	1.70	13.58	1	1	1	1	1.008E7	81	9.1	8.84
P82719	Putative defensin-like protein 148	15.03	56.79	1	4	4	11	3.728E7	81	9.2	8.68
Q1PDG8	EMBRYO SURROUNDING FACTOR 1-like protein 10	9.98	24.39	1	3	3	5	1.288E7	82	9.3	7.85
P82747	Putative defensin-like protein 150	3.11	15.66	1	1	1	1	5.114E6	83	9.3	8.84

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q00762	Tapetum-specific protein A9	14.84	69.23	2	2	2	4	2.123E7	91	9.3	6.24
Q8GW48	At4g15810	20.99	16.28	1	1	1	6	2.107E7	86	9.3	4.48
O22867	Defensin-like protein 194	2.73	20.48	1	1	1	1	2.848E6	83	9.4	8.10
Q41190	Glycine-rich protein (Fragment)	28.81	25.00	1	1	1	5	1.574E7	112	9.5	5.02
P82749	Putative defensin-like protein 142	18.69	15.29	1	1	1	7	1.358E8	85	9.6	9.13
O22866	Defensin-like protein 193	5.65	15.56	1	1	1	2	6.144E5	90	9.7	8.98
Q8LFM2	Gibberellin-regulated protein 10	21.62	26.97	1	3	3	10	1.617E7	89	9.7	8.60
Q8GT78	Protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	22.98	29.47	5	2	2	9	1.048E7	95	9.8	4.73
P82635	Putative defensin-like protein 238	4.89	28.74	1	1	1	1	1.819E7	87	9.9	9.38
P82622	Putative defensin-like protein 228	31.32	25.00	1	1	1	9	2.821E7	88	9.9	8.43
Q2V3K0	Putative defensin-like protein 169	6.10	11.49	1	1	1	3	8.753E6	87	9.9	9.26
P82716	Defensin-like protein 147	41.70	30.34	1	4	4	23	4.074E7	89	10.0	9.28
Q42039	Tonoplast intrinsic protein (Fragment)	0.00	15.38	1	1	1	1		91	10.0	5.36
Q9LTK4	At5g52160	14.79	25.00	1	1	1	4	2.967E6	96	10.1	6.32
P82644	Putative defensin-like protein 231	115.23	20.69	1	2	2	34	8.763E7	87	10.1	8.18
F4IQJ4	Gibberellin-regulated protein 11	14.09	14.89	1	1	1	5	2.019E6	94	10.1	8.25
Q9T0E3	Defensin-like protein 151	0.00	11.46	1	1	1	1		96	10.2	8.81
P82643	Putative defensin-like protein 230	7.47	24.72	1	2	2	5	9.244E7	89	10.3	8.31
P57752	Acyl-CoA-binding domain-containing protein 6	38.29	13.04	1	1	1	12		92	10.4	5.24
P82621	Defensin-like protein 226	2.05	8.70	1	1	1	1	1.351E7	92	10.5	8.48
Q9C9N7	At1g66850	383.70	60.78	1	7	7	96	2.875E8	102	10.5	7.87
Q42574	Oleosin	46.68	23.58	2	3	3	16	1.764E8	106	10.7	9.58
P82646	Defensin-like protein 229	17.29	8.60	1	1	1	11	4.640E6	93	10.7	8.88
P82642	Defensin-like protein 232	3.36	15.79	1	1	1	2	1.938E6	95	10.9	7.50
P82639	Putative defensin-like protein 236	23.49	36.46	1	4	4	11	6.662E6	96	10.9	9.13
P55852	Small ubiquitin-related modifier 1	17.39	28.00	1	2	3	5	4.627E7	100	11.0	5.10
Q9SK39	Probable steroid-binding protein 3	0.00	23.00	1	1	1	1	3.251E6	100	11.0	4.88
A8MR90	Protease inhibitor/seed storage/LTP family protein	3.17	20.83	1	1	1	2	4.173E6	96	11.0	8.95
P82638	Defensin-like protein 241	30.52	10.20	1	1	1	16	1.870E7	98	11.1	8.90
P82641	Putative defensin-like protein 233	7.53	7.14	1	1	1	6		98	11.2	7.77
Q945Q1	Cysteine proteinase inhibitor 1	2.41	26.73	1	1	1	2	2.657E6	101	11.2	5.22
A8MQR9	Protease inhibitor/seed storage/LTP family protein	10.35	12.00	1	1	1	5	2.524E7	100	11.3	8.94
Q9LLR6	Non-specific lipid-transfer protein 4	11.05	28.57	1	1	1	3	8.428E6	112	11.4	8.75
Q9T0H6	Protein unfertilized embryo sac 15	33.84	34.86	1	3	3	21	5.112E7	109	11.6	7.34
Q84TF9	At5g03370	0.00	16.04	4	1	1	1	3.517E6	106	11.6	7.14
Q9FLP6	Small ubiquitin-related modifier 2	10.07	27.18	2	2	3	3	3.485E7	103	11.6	5.55
Q9FNE2	Glutaredoxin-C2	10.10	21.62	2	1	1	2	7.033E6	111	11.7	7.24
A8MQL3	ECA1-like gametogenesis related family protein	3.42	12.50	1	1	1	1	1.615E7	104	11.8	9.07
Q9ZUX4	Uncharacterized protein At2g27730, mitochondrial	3.26	13.27	1	1	1	1	4.737E6	113	11.9	9.64
O04254	At4g02140	3.68	35.71	1	1	1	1		112	12.1	6.32
O23138	Cytochrome c-1	7.84	27.19	3	3	3	5	1.844E6	114	12.4	9.31
Q2V3C1	Non-specific lipid-transfer protein 11	54.23	42.86	1	4	4	21	3.715E7	119	12.6	8.48
Q84WD6	Putative uncharacterized protein At1g29140 (Fragment)	2.20	8.70	6	1	1	1	3.003E7	115	12.7	4.64
O65920	At2g19000	43.74	21.60	2	3	3	20	6.592E7	125	12.7	8.12

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9FMN0	Putative uncharacterized protein At5g42890	8.82	21.95	1	2	2	3	6.494E6	123	13.6	9.20
Q8L8T2	Glutaredoxin-C1	3.09	14.40	1	1	1	1	2.507E6	125	13.6	5.48
Q3E6Q3	Ras-related small GTP-binding protein	6.28	10.66	4	1	1	2	7.968E5	122	13.8	5.07
O23430	Putative uncharacterized protein AT4g15790	2.48	8.94	3	1	1	1	4.988E6	123	13.9	4.91
Q56YZ1	Putative aspartic protease	1.92	9.38	3	1	1	2	2.327E7	128	14.0	4.81
Q42449	Profilin-1	6.71	14.50	1	1	1	2	1.505E6	131	14.3	4.82
Q42512	Protein COLD-REGULATED 15A, chloroplastic	8.22	28.78	3	5	5	7	1.245E7	139	14.6	7.24
Q38879	Thioredoxin H2	11.21	24.81	2	2	2	3	1.584E7	133	14.7	6.01
Q8LFQ6	Glutaredoxin-C4	5.08	16.30	1	2	2	3	5.874E6	135	14.8	5.96
Q9C8W7	At1g71950	36.61	9.56	2	1	1	14	2.151E8	136	14.8	6.15
Q9SV91	At4g10300	7.25	17.91	2	1	1	2	8.892E6	134	14.9	8.68
Q42342	Cytochrome b5 isoform E	5.63	32.84	1	2	2	2	1.641E7	134	15.1	5.33
Q94CG2	Bet1-like SNARE 1-2	3.24	10.77	1	1	1	1	2.218E6	130	15.2	9.26
Q9FDW8	Cytochrome b5 isoform A	6.96	13.33	1	1	1	2	4.405E6	135	15.2	4.68
Q9LY07	Glycine-rich protein 20	39.06	24.18	1	1	1	9	8.037E7	153	15.4	10.49
Q9FLY6	Gb|AAD30637.1	157.98	72.73	1	14	14	81	1.481E8	132	15.6	9.57
O65387	F12F1.21 protein	4.94	5.00	2	1	1	3		140	15.6	9.44
Q9FNP8	40S ribosomal protein S19-3	3.78	5.59	3	1	1	2	4.579E6	143	15.7	10.21
Q8L5T9	Cysteine proteinase inhibitor 2	54.35	34.69	1	6	6	26	2.364E8	147	16.1	10.07
Q9SAA0	At1g11763	11.47	17.52	1	1	1	3	4.300E6	137	16.2	9.26
Q1G3D1	Putative uncharacterized protein	4.15	17.52	3	3	3	3	3.686E6	137	16.2	9.67
F4HWK0	Self-incompatibility protein S1 family protein	7.18	24.82	2	2	2	2	1.264E7	137	16.3	9.11
Q38935	Peptidyl-prolyl cis-trans isomerase FKBP15-1	21.65	41.18	2	4	4	6	5.329E7	153	16.3	8.37
Q03251	Glycine-rich RNA-binding protein 8	7.06	20.71	5	2	2	2		169	16.6	5.68
B3H4Y0	D-galactoside/L-rhamnose binding SUEL lectin protein	95.16	25.00	1	6	6	44	2.587E7	152	16.8	8.76
Q03250	Glycine-rich RNA-binding protein 7	111.49	39.77	4	5	5	39	2.341E7	176	16.9	6.15
Q9SCQ3	D-galactoside/L-rhamnose binding SUEL lectin protein	53.93	18.06	3	3	3	18	3.504E7	155	17.0	8.25
F4K817	Glycine-rich protein 14	18.23	23.20	8	1	1	4	2.342E8	181	17.1	11.30
Q8RUC6	Ubiquitin-NEDD8-like protein RUB2	33.24	21.43	40	2	4	25	1.301E8	154	17.1	6.06
Q570J3	Adenylosuccinate synthetase (Fragment)	3.02	11.80	3	1	1	1	1.271E7	161	17.7	6.57
Q93YP0	Ubiquitin-conjugating enzyme E2 variant 1A	1.90	7.59	1	1	1	1	1.045E7	158	17.8	5.24
Q9LE56	At1g18280	23.79	15.56	2	3	3	11	7.619E6	180	18.1	7.84
Q42406	Peptidyl-prolyl cis-trans isomerase CYP18-4	6.41	6.98	1	1	1	3	6.531E6	172	18.4	8.69
B3H6K7	D-galactoside/L-rhamnose binding SUEL lectin protein	57.93	41.82	1	6	6	14	3.860E8	165	18.4	9.82
O04922	Probable glutathione peroxidase 2	5.37	13.02	1	2	2	2	1.394E6	169	18.9	5.72
Q9FNM0	Putative uncharacterized protein At5g46020	4.27	26.22	1	2	2	2	1.789E6	164	19.0	7.27
A0A097PSQ3	AT4G29520-like protein (Fragment)	2.86	8.93	3	1	1	1	1.148E7	168	19.0	5.53
F4JGJ8	Uncharacterized protein	2.47	14.97	1	1	1	1	3.472E7	167	19.0	4.46
Q38896	Cold shock domain-containing protein 4	37.94	35.82	1	4	4	8	9.236E6	201	19.1	6.76
Q41188	Cold shock protein 2	6.66	13.79	1	2	2	2	9.123E6	203	19.1	5.92
Q570C6	Putative uncharacterized protein At3g62370 (Fragment)	0.00	8.94	3	1	1	1	8.716E5	179	19.9	5.99
Q9ZPQ7	At2g03740	73.61	31.22	4	6	7	27	3.670E7	189	20.0	9.32
Q07488	Blue copper protein	2.68	8.16	1	1	1	1	4.401E5	196	20.0	4.82
Q9SEU8	Thioredoxin M2, chloroplastic	4.94	9.68	1	1	1	1	5.758E6	186	20.3	9.22

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q9SI55	At2g03850	38.90	12.04	3	2	3	22	5.260E7	191	20.6	9.19
P0DH91	ADP-ribosylation factor 2-B	14.10	7.18	6	2	2	6	8.135E5	181	20.6	6.95
Q93WI3	Glycine-rich protein GRP18	28.87	20.83	6	2	2	8	2.577E7	216	20.6	9.91
P42763	Dehydrin ERD14	14.09	7.03	1	1	1	6	1.105E7	185	20.8	5.48
Q93YN0	Protein disulfide-isomerase SCO2	2.47	8.56	1	1	1	1		187	20.8	7.83
Q94BU3	T29M8.1/T29M8.1	3.39	4.59	5	1	1	3	2.626E6	196	21.5	9.99
Q9FJR2	Plant invertase/pectin methylesterase inhibitor domain-containing protein	8.80	27.81	1	4	4	5	3.639E6	187	21.6	9.10
Q8H7C0	Putative uncharacterized protein (Fragment)	4.22	10.36	3	1	1	1	2.331E7	193	21.6	7.94
O04834	GTP-binding protein SAR1A	2.21	12.44	4	2	2	2	4.176E6	193	22.0	7.53
Q9FLS1	Oleosin family protein	27.64	13.04	1	2	2	14	1.108E7	230	22.1	10.30
Q9FPJ4	Ras-related protein RABD2b	5.97	7.92	2	1	1	2	4.900E6	202	22.3	5.43
Q9SSM4	ESCRT-related protein CHMP1B	0.00	3.94	1	1	1	1	3.040E6	203	22.7	7.20
Q9FRL8	Glutathione S-transferase DHAR2	2.47	6.57	1	1	1	2	2.761E6	213	23.4	6.14
B9DHZ4	AT1G11755 protein (Fragment)	5.80	3.77	2	1	1	4	1.442E7	212	23.8	6.21
B3H7A9	Probable peroxygenase 7	13.40	6.19	1	2	2	6	1.092E8	210	23.8	9.61
Q93V66	Protein FATTY ACID EXPORT 1, chloroplastic	10.46	11.50	1	1	1	3	2.809E6	226	24.3	9.03
P59230	60S ribosomal protein L10a-2	1.73	3.70	3	1	1	1	2.693E6	216	24.4	9.88
A6XI99	Ubiquitin (Fragment)	43.61	47.95	19	1	3	21	8.836E7	219	24.6	6.83
A0JQ12	At1g76020	0.00	9.33	2	1	1	1	6.453E6	225	25.2	8.16
Q9SKI2	Vacuolar protein sorting-associated protein 2 homolog 1	13.76	10.67	1	1	1	3	5.215E6	225	25.3	5.63
O24468	Thaumatin-like protein	16.49	5.31	2	1	1	5	2.852E7	245	25.4	5.19
Q9C9P9	At1g75030	11.37	34.96	3	3	3	3	4.626E6	246	25.4	5.19
Q94BT2	Auxin-induced in root cultures protein 12	2.64	3.57	1	1	1	1	2.976E7	252	25.6	7.90
P42815	Ribonuclease 3	52.55	56.76	1	7	7	17	4.954E7	222	25.6	6.14
F4I699	Uncharacterized protein	16.80	5.48	4	2	2	9	4.587E7	219	25.8	9.86
Q8RXH8	Chloroplast-targeted copper chaperone protein	0.00	3.64	3	1	1	1		247	26.1	9.63
Q8H0X6	Cysteine proteinase inhibitor 6	5.70	6.84	1	1	1	2	2.924E7	234	26.3	6.27
F4I6A3	Uncharacterized protein	10.60	7.56	3	2	2	3	3.412E7	225	26.5	9.55
Q8LFH5	Bidirectional sugar transporter SWEET8	3.00	5.86	1	1	1	2	8.969E5	239	26.9	9.23
Q9LFD5	Binding partner of ACD11 1	80.64	30.12	4	5	5	31	1.705E7	259	27.2	5.63
O23715	Proteasome subunit alpha type-3	16.91	3.61	1	1	1	9	1.699E7	249	27.4	6.32
Q9XIL1	At2g15780/F19G14.22	2.65	5.06	1	1	1	1	1.443E6	257	27.4	9.51
F4K5V1	NDR1/HIN1-like 25	1.70	2.02	9	1	1	1	9.986E6	248	27.9	9.85
Q9SCK3	At3g49560	11.05	6.51	1	1	1	3	1.152E7	261	28.0	9.50
F4I1C1	14-3-3-like protein GF14 epsilon	0.00	3.98	4	1	1	1	6.496E6	251	28.6	4.92
C0Z2H0	AT5G60390 protein	6.98	8.68	10	2	2	3	2.799E5	265	28.8	9.55
Q9LRK8	Cysteine-rich repeat secretory protein 29	44.10	49.22	8	9	10	14	2.232E8	256	28.9	8.73
Q9LRK4	Putative cysteine-rich repeat secretory protein 33	3.62	3.56	6	1	2	2	2.985E8	253	29.0	8.75
Q01525	14-3-3-like protein GF14 omega	4.63	4.63	1	1	1	2	1.949E8	259	29.1	4.79
Q8VZU9	Derlin-1	24.42	9.77	1	1	1	5	9.984E5	266	29.2	9.95
F4IL53	Protein disulfide-isomerase like 2-1	49.58	21.80	5	7	7	19	4.682E8	266	29.2	5.76
Q9M2I5	Putative cysteine-rich repeat secretory protein 61	16.80	22.09	1	4	4	6	5.094E7	258	29.5	9.01
Q9LRL4	Putative cysteine-rich repeat secretory protein 24	16.47	15.65	2	3	3	6	1.496E8	262	29.6	8.69
Q9LRK2	Putative cysteine-rich repeat secretory protein 35	1.71	11.24	4	1	3	5	2.038E8	258	29.6	9.00

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
O23717	Proteasome subunit beta type-5-A	6.41	8.39	1	1	1	3		274	29.6	6.43
Q96300	14-3-3-like protein GF14 nu	3.29	6.42	1	1	1	1	5.695E7	265	29.8	4.82
O22782	Putative uncharacterized protein At2g33300	0.00	2.26	3	1	1	1		265	29.9	9.88
P42643	14-3-3-like protein GF14 chi	10.63	5.62	2	1	1	4	6.368E7	267	29.9	4.81
F4K975	Sec14p-like phosphatidylinositol transfer family protein	2.29	6.46	2	1	1	4	1.264E6	263	30.1	8.79
P42645	14-3-3-like protein GF14 upsilon	19.06	5.60	1	1	1	7	6.848E7	268	30.2	4.81
Q9ZUD1	F5O8.19 protein	0.00	6.82	1	1	1	3	1.649E6	264	30.3	9.45
Q9LRL1	Cysteine-rich repeat secretory protein 26	70.87	28.03	5	8	10	21	2.136E8	264	30.3	8.91
K9M8T1	OTU-containing deubiquitinating enzyme 4 isoform iia	11.97	13.14	3	2	2	4	3.206E7	274	30.8	8.57
B9DGT0	Lactoylglutathione lyase	2.85	3.89	4	1	1	1	5.168E7	283	32.0	5.27
F4K3K4	Oleosin	50.87	18.71	2	4	4	22	2.738E7	294	32.3	8.66
Q8LAN3	Probable prolyl 4-hydroxylase 4	2.54	8.39	1	1	1	1	3.885E6	298	33.0	7.15
P19172	Acidic endochitinase	18.20	20.53	1	3	3	5	2.031E7	302	33.1	9.01
P92976	Protein STRICTOSIDINE SYNTHASE-LIKE 11	5.59	5.47	1	1	1	2	2.008E7	329	34.6	9.64
Q9LZS8	GDSL esterase/lipase At5g03600	0.00	6.21	1	1	1	1	2.398E7	322	35.7	9.17
Q8L7U0	AT3g03330/T21P5_25	2.11	3.96	2	1	1	1	1.562E6	328	35.8	9.22
Q5Q0E3	Putative uncharacterized protein	2.23	5.96	5	1	1	1	1.863E6	319	36.1	5.54
F4KCF2	NAD(P)-binding Rossmann-fold superfamily protein	2.08	4.23	4	1	1	1	9.719E6	331	36.2	8.40
Q940G5	Aldose 1-epimerase family protein	11.83	29.25	4	5	5	7	4.023E7	318	36.2	9.66
A8MSC5	AT2G47470 protein	40.53	13.43	6	5	5	14	1.518E7	335	36.5	5.67
Q8VZV6	Esterase/lipase/thioesterase family protein	7.42	3.94	1	1	1	2	3.063E7	330	36.6	6.79
Q9ZVH7	Protein FATTY ACID EXPORT 3, chloroplastic	3.96	5.97	1	1	1	1	6.214E5	335	36.7	7.53
F4J1V2	Chaperone protein dnaJ 3	0.00	4.37	3	1	1	1	5.339E5	343	37.7	8.18
Q9LNC1	Cysteine proteinases superfamily protein	40.06	26.53	1	6	6	9	2.678E7	343	37.7	7.58
Q0WUV7	GDSL esterase/lipase EXL4	1.70	2.33	1	1	1	1	7.796E6	343	37.9	9.80
F4JQ50	Lysophosphatidic acid acyltransferase	12.22	7.60	2	1	1	4	6.392E5	342	38.0	9.55
Q9LF11	Putative uncharacterized protein T21H19_90	0.00	3.04	2	1	1	1		329	38.0	7.17
Q8LA49	Globulin-like protein	3.17	6.18	2	1	1	1	2.423E6	356	38.3	6.18
Q8LCU2	Putative uncharacterized protein	2.80	6.65	3	1	1	1	2.148E6	346	38.7	5.69
Q9SJS8	DNAJ heat shock N-terminal domain-containing protein	4.38	5.20	1	1	1	2	5.254E7	346	38.8	7.99
Q9LZK5	DnaJ protein ERDJ3B	3.88	4.62	1	1	1	1	2.341E6	346	39.2	6.29
F4I0R0	GDSL esterase/lipase EXL5	3.91	5.10	2	1	1	1	4.952E5	353	39.2	9.57
Q8GXU8	1-acyl-sn-glycerol-3-phosphate acyltransferase 1, chloroplastic	8.51	5.34	1	1	1	2	8.879E5	356	39.4	9.85
Q9FXA0	F14J22.5 protein	54.75	30.64	3	7	7	15	1.280E7	359	39.8	8.12
Q949V0	AT5G51180 protein	5.91	3.36	2	1	1	2	1.743E6	357	40.0	5.63
O82291	Probable plastid-lipid-associated protein 3, chloroplastic	2.06	2.93	1	1	1	1	3.042E6	376	40.5	4.55
Q9SNY3	GDP-mannose 4,6 dehydratase 1	2.25	2.77	1	1	1	1		361	40.8	7.11
F4J8V9	Actin 2	19.85	5.66	14	2	2	9	6.095E6	371	41.2	5.69
Q9SUL1	AT4g16190/dl4135w	3.72	6.17	1	1	1	1	2.443E6	373	41.2	6.96
Q9LVD4	Oxysterol-binding protein-related protein 4C	13.16	3.17	1	1	1	6	1.765E6	379	42.9	9.06
Q8W4H8	Inactive GDSL esterase/lipase-like protein 23	4.86	3.11	1	1	1	2	1.543E7	386	43.1	8.07
Q8L7C5	Putative uncharacterized protein At5g39570	1.91	2.59	1	1	1	1		386	43.4	10.70
Q8L7R3	Lysophospholipid acyltransferase LPEAT1	9.61	3.52	1	1	1	3	2.383E6	398	44.7	6.01
F4IBV4	Glycosyltransferase	2.53	4.22	3	1	1	1	4.467E6	403	45.0	7.15

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
O64517	Metacaspase-4	7.53	2.63	1	1	1	3	1.682E6	418	45.5	4.82
B3LF50	At1g11770	2.17	5.67	5	2	2	2	4.591E6	406	45.5	9.39
P49063	Exopolygalacturonase clone GBGA483	5.53	5.18	5	3	3	5	1.162E7	444	45.6	8.31
Q944P0	Equilibrative nucleotide transporter 7	4.69	6.00	1	1	1	2	5.204E6	417	45.7	8.18
F4I5X3	Alpha/beta-Hydrolases superfamily protein	2.60	3.69	2	1	1	1	1.750E7	406	45.9	5.91
A8MS20	F-box protein At2g43440	1.74	2.99	2	1	1	1	2.831E5	401	46.2	8.25
Q9M0Y3	Equilibrative nucleotide transporter 3	0.00	6.22	1	1	1	1	1.460E6	418	46.2	8.65
B9DFF8	AT4G14960 protein	3.27	3.51	5	1	1	1		427	47.2	8.09
Q93VB4	MYND type zinc finger and programmed cell death 2 C-terminal domain-containing protein	0.00	1.20	2	1	1	1	3.588E6	418	47.5	4.92
Q0WP75	Kinesin like protein (Fragment)	0.00	1.93	2	1	1	1		415	47.9	6.58
A8MQE5	Putative mitochondrial-processing peptidase subunit alpha-1	2.38	2.22	3	1	1	1		451	48.6	6.00
Q9SD56	Hypersensitivity related-like protein, Nicotiana tabacum, X95343	2.04	3.13	2	1	1	1	6.356E6	447	49.3	6.65
B9DI49	AT1G15500 protein (Fragment)	3.81	5.71	2	1	1	1	1.151E6	455	49.9	9.26
Q9T081	UDP-glycosyltransferase 79B3	2.88	6.62	1	1	1	1	2.981E7	453	50.3	5.82
Q9LFZ4	F20N2.17	2.21	2.99	1	1	1	1	7.059E6	435	50.5	8.65
Q9LXI4	Purple acid phosphatase 21	32.26	30.43	1	6	6	16	1.010E7	437	50.6	9.01
P46312	Omega-6 fatty acid desaturase, chloroplastic	8.66	3.79	1	1	1	3	6.113E5	448	51.2	8.84
O23254	Serine hydroxymethyltransferase 4	1.76	1.91	1	1	1	1	7.524E6	471	51.7	7.23
Q93Y40	Oxysterol-binding protein-related protein 3C	3.99	3.28	1	1	1	2	2.433E6	457	51.9	5.30
Q8VYN0	Putative uncharacterized protein At1g65985	3.00	5.92	2	1	1	1		456	53.1	6.86
Q9LY09	Oleosin GRP-17	56.93	24.86	5	8	8	12	9.401E7	543	53.2	10.33
P92940	CaLB protein	5.02	3.25	2	2	2	2	3.074E6	493	53.4	7.88
Q6NMN6	Probable sphingolipid transporter spinster homolog 1	18.37	5.49	1	2	2	6	1.510E7	492	53.5	8.69
F4K4J2	Lipase class 3-like protein	0.00	1.50	1	1	1	1	6.390E7	467	54.3	7.08
F4JGD2	Aspartic proteinase A3	0.00	1.39	2	1	1	1		504	55.1	7.34
Q56WF8	Serine carboxypeptidase-like 48	3.12	2.75	1	1	1	1	2.420E6	510	56.9	5.50
P32826	Serine carboxypeptidase-like 49	23.98	4.26	1	3	3	28	6.755E7	516	57.3	5.38
Q67XX3	F-box protein At5g06550	0.00	5.58	1	1	1	1	3.220E7	502	57.4	5.55
P0C7R3	Pentatricopeptide repeat-containing protein At1g64583, mitochondrial	1.64	1.37	1	1	1	1	3.076E5	512	57.9	7.52
Q6Q151	Peptidyl-prolyl cis-trans isomerase CYP59	0.00	4.35	1	1	1	1	1.282E7	506	58.8	6.25
O64743	FAD-binding and BBE domain-containing protein	3.43	3.57	1	1	1	1	5.576E6	532	59.6	6.42
Q9CAT6	Organic cation/carnitine transporter 1	6.98	4.08	1	1	1	7	1.612E6	539	59.6	8.46
F4I340	Putative adipose-regulatory protein (Seipin)	5.65	6.27	1	1	1	1	2.418E7	526	59.6	5.19
Q8GYT8	Putative uncharacterized protein At2g28150	2.20	2.04	2	1	1	2		540	60.0	7.24
Q8VY11	DNA gyrase B3	0.00	3.11	1	1	1	1	9.593E6	546	60.9	8.00
Q9SKR2	Synaptotagmin-1	9.23	4.81	2	1	1	2		541	61.7	7.56
Q9LP92	Putative uncharacterized protein At1g36990	0.00	2.58	1	1	1	1	2.968E6	581	61.7	8.54
Q9SR96	DnaJ protein ERDJ3A	4.17	2.80	1	1	1	1	2.783E6	572	62.5	9.50
Q9LH95	Genomic DNA, chromosome 3, BAC clone: T19N8	13.56	2.80	1	1	1	3	3.651E6	608	62.7	9.38
Q8L706	Synaptotagmin-5	4.80	3.39	1	1	1	1	2.107E6	560	62.9	5.87
Q5MFV6	Probable pectinesterase/pectinesterase inhibitor VGDH2	3.55	2.89	1	1	1	1	4.685E5	588	62.9	8.54
Q9SF04	F26K24.24 protein	2.12	2.27	1	1	1	1	1.805E6	572	64.2	4.70

Accession	Description	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
Q8LA13	DEAD-box ATP-dependent RNA helicase 11	11.33	1.63	1	1	1	6	7.190E5	612	66.0	6.68
Q9XID4	Auxin response factor 12	0.00	3.54	1	1	1	1	8.610E6	593	67.2	6.71
F4J116	Uncharacterized protein	1.68	1.51	2	1	1	1		596	67.5	8.65
F4JBE2	Terminal EAR1-like 1	1.96	1.14	2	1	1	1		615	69.4	8.24
Q84MA9	Inactive leucine-rich repeat receptor-like serine/threonine-protein kinase At1g60630	5.14	0.92	1	1	1	3		652	72.3	8.65
Q9C9U3	Alpha-dioxygenase 2	5.73	1.74	1	1	1	2	2.295E5	631	72.4	6.79
Q9SGA1	F1C9.19 protein	1.77	4.42	2	2	2	3	1.229E7	656	73.4	8.41
Q0WVS3	Trichohyalin like protein (Fragment)	0.00	2.43	3	1	1	1		699	77.7	7.11
Q9ZU46	Leucine-rich repeat receptor-like protein kinase	1.87	1.12	1	1	1	1		716	78.3	6.11
O65351	Subtilisin-like protease SBT1.7	28.67	13.34	1	3	3	7	7.296E6	757	79.4	6.35
Q9SCV8	Beta-galactosidase 4	2.36	1.66	6	1	1	1	4.044E5	724	80.5	8.68
A0A097PJP3	Minichromosome maintenance 5 protein (Fragment)	0.00	1.24	3	1	1	2	2.698E7	723	80.6	7.44
O23337	Pentatricopeptide repeat-containing protein At4g14820	0.00	3.19	1	1	1	1		722	82.1	6.46
Q304B9	Neutral ceramidase	32.78	11.10	2	4	4	8	3.446E7	757	83.2	8.53
Q9SRV5	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase 2	2.80	5.23	5	2	2	3	1.004E7	765	84.5	6.51
Q710E8	Origin of replication complex subunit 1A	0.00	0.87	2	1	1	1	1.143E6	809	91.8	7.47
Q9C622	Receptor serine/threonine kinase PR5K, putative	0.00	0.57	1	1	1	1	1.589E6	876	98.1	6.05
Q9LTY0	Putative uncharacterized protein	16.83	3.23	4	1	1	5	2.316E7	928	103.8	6.16
Q9M2T1	AP3-complex subunit beta-A	4.39	1.11	2	1	1	2	2.041E6	987	108.6	5.48
F4J4L4	Uncharacterized protein	0.00	2.25	1	1	1	1		1023	114.9	7.50
Q6IMG0	GRP21	5.07	3.35	1	1	1	1	4.444E6	1193	115.4	9.50
Q15ED8	PHYTOCHROME C (Fragment)	0.00	0.66	22	1	1	1	2.634E5	1054	117.1	6.23
Q9LHP4	Receptor-like protein kinase 2	2.48	3.33	1	1	1	1	6.652E6	1141	124.4	5.45
Q9FHF0	Disease resistance protein-like	0.00	0.60	1	1	1	1	1.159E5	1165	132.3	8.10
Q9ZQK0	Putative retroelement pol polyprotein	0.00	1.14	1	1	1	1		1664	189.9	7.83

Table S6.2 | Statistically validated* hits from mass spectrometry analysis of proteins obtained from *Brassica oleracea* pollen coat. Score: the sum of the scores of the individual peptides; Coverage: the percentage of the protein sequence covered by identified peptides; Proteins: the number of identified proteins in the protein group of a master protein; Unique peptides: the number of peptide sequences unique to a protein group; Peptides: the number of distinct peptide sequences in the protein group; PSMs: the number of peptide-spectrum match (PSM) for the protein, including those redundantly identified; Area: the average area of the three unique peptides with the largest peak area; pI: theoretically calculated isoelectric point, which is the pH at which a particular molecule carries no net electrical charge. The PSMs were statistically validated to avoid false positives by using the False discovery rate (FDR)-controlling procedure (for details see 2.9.4).

Accession	Homologous protein in Brassicaceae and sequence similarity	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A0D2ZR09	-	21.69	64.81	1	1	3	8	7.458E9	54	5.8	8.87
A0A0D3A650	EMBRYO SURROUNDING FACTOR 1-like protein 9 (*Arabidopsis thaliana*) 34.2%	25.69	39.39	1	3	3	10	3.600E9	66	7.5	8.90
A0A0D3B5U8	Protein RALF-like 36 (*Arabidopsis thaliana*) 47.7%	12.81	26.03	1	2	2	4	2.566E8	73	7.6	8.12
A0A0D3DAB5	Defensin-like protein 10 (*Arabidopsis thaliana*) 71.2%	20.73	27.40	1	1	1	4	3.325E8	73	7.7	8.32
A0A0D3E4B8	Defensin-like protein 35 (*Arabidopsis thaliana*) 71.1%	1.89	10.67	1	1	1	1	8.987E5	75	8.0	8.31
A0A0D3BF83	Protein BP4A/C (*Brassica napus*) 89%	51.22	60.27	2	7	8	19	5.224E7	73	8.1	8.82
A0A0D3BI08	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (*Arabidopsis thaliana*)	4.36	13.16	2	2	2	2	3.157E7	76	8.2	9.25
A0A0D3AJX4	BnaC02g08710D protein	367.45	76.71	2	9	9	141	4.202E10	73	8.3	9.29
A0A0D3ED66	Putative defensin-like protein 274 (*Arabidopsis thaliana*) 77.2%	3.64	33.77	1	1	1	1	2.855E7	77	8.3	6.71
A0A0D3BR76	Defensin-like protein 10 (*Arabidopsis thaliana*) 42.3%	57.00	34.25	1	2	2	14	2.700E9	73	8.6	6.98
A0A0D3C583	Putative defensin-like protein 62 (*Arabidopsis thaliana*) 42%	4.78	34.15	2	1	1	1	1.482E8	82	8.8	8.32
A0A0D3BQ22	Putative defensin-like protein 15 (*Arabidopsis thaliana*) 46.3%	5.52	8.75	1	1	1	2	1.389E8	80	9.1	8.85
A0A0D3DZ37	PCP-A1	55.69	51.85	1	6	6	27	1.852E10	81	9.2	9.11
A0A0D3BQ21	Putative defensin-like protein 15 (*Arabidopsis thaliana*) 42.5%	7.89	11.25	1	1	1	4	8.770E8	80	9.2	8.66
A0A0D3C3T8	Defensin-like protein 151 (*Arabidopsis thaliana*) 34.4%	30.94	38.10	1	4	4	13	1.430E10	84	9.3	8.97
A0A0D3BZ12	Defensin-like protein 147 (*Arabidopsis thaliana*) 42.2%	189.18	54.76	1	9	9	57	3.506E10	84	9.4	9.00
A0A0D3DIU7	Pollen coat protein (*Brassica oleracea*) PCP-1 96.4%	84.62	46.43	1	5	5	30	1.714E10	84	9.4	9.32
A0A0D3CTX4	Protein BP4C (*Brassica napus*) 48.8%	51.00	48.19	1	5	5	16	8.241E9	83	9.5	8.98
A0A0D3BZ11	Pollen coat protein (*Brassica oleracea*) PCP-2 51.2%	54.48	29.89	1	3	3	22	1.237E10	87	9.5	8.97
A0A0D2ZR06	Putative defensin-like protein 145 (*Arabidopsis thaliana*) 33%	65.40	63.64	1	3	6	16	4.539E9	88	9.5	8.79
A0A0D3DZ39	BnaC08g43580D protein (*Brassica napus*)	19.47	41.67	1	4	4	10	6.828E9	84	9.5	8.66
A0A0D3CPP4	PCP-1 family (*Brassica campestris*) 63.1%	10.40	40.70	1	3	3	4	1.117E7	86	9.5	9.16
A0A0D3CPP6	PCP-1 family (*Brassica campestris*) 59.5%	2.04	9.30	1	1	1	1	4.276E5	86	9.6	9.45
A0A0D3BJE6	F-box/kelch-repeat protein At4g23580 (*Arabidopsis thaliana*) 70.8%	2.56	21.69	1	1	1	1		83	9.6	9.14
A0A0D3DBH4	Protease inhibitor/seed storage/LTP family protein (*Arabidopsis thaliana*) 38.1%	10.04	30.00	1	1	1	3	6.071E7	90	9.7	5.81
A0A0D3CYH8	Lipid transfer protein (*Brassica campestris*) 82.6%	209.21	65.26	1	2	8	63	3.439E10	95	9.8	8.50
A0A0D3BPK4	Defensin-like protein 195 (*Arabidopsis thaliana*) 83%	18.99	38.64	2	3	3	5	1.620E8	88	9.9	7.50
A0A0D3AKV6	Gibberellin-regulated family protein (*Arabidopsis lyrata*) 87.6%	26.75	31.46	2	2	2	10	1.159E10	89	9.9	8.47
A0A0D3CUV4	Allyl alcohol dehydrogenase-like protein (*Arabidopsis thaliana*) 91.8%	14.84	34.78	1	1	3	5	4.359E7	92	9.9	4.93
A0A0D3B3V3	At5g52160 (*Arabidopsis thaliana*) 77.9%	9.80	18.56	1	1	1	3	7.720E7	97	9.9	6.93
A0A0D3DBH5	Protease inhibitor/seed storage/LTP family protein (*Arabidopsis thaliana*) 37.4%	54.92	71.28	1	5	5	17	1.454E8	94	10.1	7.58
A0A0D3A9C5	Putative defensin-like protein 38 (*Arabidopsis thaliana*) 32.9%	19.35	20.88	1	2	2	5	1.582E9	91	10.1	8.95
A0A0D3B401	SCRL1 protein (*Brassica oleracea*) 96.9%	145.05	41.30	1	6	6	43	4.387E10	92	10.2	8.06
A0A0D3CGX6	Acyl-CoA-binding protein (*Brassica napus*) 97.8%	27.77	20.65	1	2	2	6	5.415E7	92	10.2	5.22

Accession	Homologous protein in Brassicaceae and sequence similarity	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A0D3BCY3	Putative defensin-like protein 251 (*Arabidopsis thaliana*) 41.3%	11.30	35.23	1	3	3	5	1.338E9	88	10.2	8.73
A0A0D2ZWK9	Putative defensin-like protein 38 (*Arabidopsis thaliana*) 51.1%	31.01	33.33	1	5	5	12	2.251E9	90	10.3	8.48
A0A0D3CF87	Acyl-CoA-binding protein (*Brassica napus*) 88%	23.66	22.58	1	2	2	6	1.877E7	93	10.3	5.50
A0A0D3AND9	Lipid transfer protein (*Brassica campestris*) 95.3%	209.61	63.27	1	2	8	62	3.439E10	98	10.3	8.12
A0A0D3E575	Defensin-like protein 206 (*Arabidopsis thaliana*) 60.4%	5.50	28.72	1	1	1	1	2.638E8	94	10.3	7.81
A0A0D3B4Z0	-	0.00	10.64	1	1	1	1	6.128E7	94	10.5	9.01
A0A0D3CKV1	Protease inhibitor/seed storage/LTP family protein (*Arabidopsis thaliana*) 30.6%	99.40	52.69	1	7	7	28	1.408E9	93	10.5	8.98
A0A0D3EBC5	Bifunctional inhibitor/lipid-transfer protein (*Arabidopsis thaliana*) 59.6%	2.68	12.87	1	1	1	1	8.151E6	101	10.6	9.01
A0A0D3CJN0	Putative defensin-like protein 233 (*Arabidopsis thaliana*) 47.9%	11.42	32.97	1	2	2	3	2.429E8	91	10.6	8.35
A0A0D3BYT4	Putative defensin-like protein 230 (*Arabidopsis thaliana*) 42.9%	7.93	14.89	2	1	2	3	2.664E8	94	10.7	7.78
A0A0D3BD82	Putative defensin-like protein 251 (*Arabidopsis thaliana*) 39.4%	29.92	39.13	1	5	5	11	2.893E7	92	10.7	9.16
A0A0D2ZYX7	Chaperone protein dnaJ 2 (*Arabidopsis thaliana*) 96%	4.62	26.00	4	2	2	2	3.200E7	100	10.7	5.10
A0A0D3D2S0	Putative defensin-like protein 145 (*Arabidopsis thaliana*) 37.1%	28.91	50.51	1	3	5	10	7.351E9	99	10.9	9.01
A0A0D3DBN8	NADH dehydrogenase 1 α subcomplex subunit 2 (*Arabidopsis thaliana*) 85.7%	0.00	19.39	2	1	1	1		98	11.0	9.22
A0A0D3EH65	Pollen coat oleosin-glycine rich protein (*Sisymbrium irio*) 65.1%	9.31	19.44	1	2	2	3	3.058E8	108	11.1	10.64
A0A0D3B7X0	Defensin-like protein 226 (*Arabidopsis thaliana*) 38.2%	170.50	48.51	1	7	7	51	4.495E10	101	11.1	8.69
A0A0D2ZWA1	ECA1-like gametogenesis related family protein (*Arabidopsis thaliana*) 42%	20.05	54.64	1	6	6	11	4.007E8	97	11.2	8.47
A0A0D3EG13	Phytocystatin (*Brassica rapa*) 98%	7.72	15.84	1	1	1	2	9.253E6	101	11.3	5.81
A0A0D2ZUE8	Beta-galactosidase 7 (*Arabidopsis thaliana*) 80.6%	10.51	28.70	2	2	2	3	8.101E7	108	11.3	8.31
A0A0D2ZQY0	Putative defensin-like protein 233 (*Arabidopsis thaliana*) 39.4%	16.12	33.01	1	4	4	6	2.152E9	103	11.5	7.64
A0A0D3C822	LTP (*Brassica rapa*) 31.6%	192.25	69.90	3	10	10	59	4.208E10	103	11.7	8.95
A0A0D2ZX73	Non-specific lipid-transfer protein 1 (*Arabidopsis thaliana*) 74.6%	13.21	16.95	1	1	1	3	1.323E8	118	11.8	8.66
A0A0D3BF50	Peptidyl-prolyl cis-trans isomerase FKBP12 (*Arabidopsis thaliana*) 92.9%	2.51	11.61	2	1	1	1	2.050E7	112	11.9	7.12
A0A0D3C500	Non-specific lipid-transfer protein 2 (*Arabidopsis thaliana*) 73.7%	13.36	13.68	1	1	1	3	9.460E7	117	11.9	8.98
A0A0D3A219	Putative defensin-like protein 230 (*Arabidopsis thaliana*) 44.4%	26.38	24.51	1	3	3	8	3.750E9	102	11.9	8.48
A0A0D3AD58	Putative defensin-like protein 233 (*Arabidopsis thaliana*) 39.8%	20.19	35.85	1	4	4	7	6.045E8	106	12.0	7.84
A0A0D3B059	Pollen coat oleosin-glycine rich protein (*Sisymbrium irio*) 79.2%	0.00	9.48	1	1	1	1	3.638E7	116	12.1	9.32
A0A0D3BAI1	Non-specific lipid-transfer protein 6 (*Arabidopsis thaliana*) 67.5%	37.35	38.26	1	3	3	8	2.503E7	115	12.1	8.13
A0A0D3B7Y2	Cytochrome c-2 (*Arabidopsis thaliana*) 96.4%	4.29	16.96	4	2	2	2	2.579E8	112	12.3	9.32
A0A0D3BLN8	Non-specific lipid-transfer protein 11 (*Arabidopsis thaliana*) 69.2%	143.42	54.62	1	8	8	48	1.567E10	119	12.4	8.10
A0A0D3DE00	Non-specific lipid-transfer protein 12 (*Arabidopsis thaliana*) 77.3%	26.51	42.02	1	3	3	7	3.943E7	119	12.5	6.33
A0A0D3E795	Cystatin domain containing protein (*Brassica oleracea*) 98.3%	11.54	22.61	1	2	2	5	3.769E8	115	12.5	8.97
A0A0D3A2I2	Putative defensin-like protein 231 (*Arabidopsis thaliana*) 46.6%	14.79	26.85	1	2	3	8	5.243E9	108	12.6	6.92
A0A0D3A081	NAC transcription factor 59 (*Brassica napus*) 72.3%	2.44	34.51	2	1	1	1	2.802E7	113	12.7	4.70
A0A0D2ZQZ3	Putative defensin-like protein 152 (*Arabidopsis thaliana*) 28.6%	10.21	21.55	1	1	1	3	6.815E7	116	12.7	8.97
A0A0D3CX24	Protein BP4C (*Brassica napus*) 90.7%	25.96	29.31	1	4	5	9	3.916E9	116	12.9	8.53
A0A0D3E648	Sterol carrier protein 2 (SCP-2) family protein (*Arabidopsis lyrata*) 95.1%	18.17	30.08	1	3	3	6	1.398E8	123	13.5	9.35
A0A0D3DM93	Cysteine proteinase inhibitor 4 (*Arabidopsis thaliana*) 39.7%	19.49	25.44	1	3	3	7	2.453E8	114	13.5	9.51
A0A0D3AKQ7	Self-incompatibility S1 family protein (*Arabidopsis thaliana*) 39.3%	45.54	41.80	1	5	5	14	1.424E9	122	14.0	9.29
A0A0D3B127	Glutaredoxin-C8 (*Arabidopsis thaliana*) 86.1%	2.85	16.79	1	1	1	1		137	14.6	8.21
A0A0D3BYU3	AtPCP-Bγ (*Arabidopsis thaliana*) 50.7%	12.16	31.34	3	3	3	5	5.761E6	134	14.8	8.53
A0A0D2ZZ53	MYB4R1 (*Arabidopsis lyrata*) 71.8%	1.76	10.69	2	1	1	1	3.970E8	131	15.2	9.91
A0A0D3D4P9	T27D20.18 protein (*Arabidopsis thaliana*) 40.4%	2.18	19.70	1	1	1	1	1.117E8	132	15.4	9.58
A0A0D3CZK9	ARF-like small GTPase (*Brassica juncea*) 99.2%	10.44	9.56	7	2	2	4	8.160E7	136	15.6	5.30
A0A0D3CA11	At3g44100 (*Arabidopsis thaliana*) 80.3%	3.18	13.16	1	1	1	1	9.843E7	152	16.3	7.42

Accession	Homologous protein in Brassicaceae and sequence similarity	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A0D3BX40	D-galactoside/L-rhamnose binding SUEL lectin protein (*Arabidopsis thaliana*) 46.7%	103.05	45.21	1	7	7	32	1.026E10	146	16.3	8.22
A0A0D3EH61	Glycine-rich protein GRP18 (*Arabidopsis thaliana*) 46.2%	30.42	38.60	1	3	3	8	4.209E8	171	16.7	10.39
A0A0D3CSD2	tRNA (Cytosine-5-)-methyltransferase (*Gossypium arboreum*) 51.2%	10.10	17.95	1	2	2	5	1.906E7	156	16.7	7.24
A0A0D3CFL3	-	1.70	4.96	4	1	1	1		141	16.8	9.86
A0A0D3BYC3	D-galactoside/L-rhamnose binding SUEL lectin protein (*Arabidopsis thaliana*) 45.7%	61.20	34.42	1	3	5	18	2.186E10	154	17.2	8.28
A0A0D3BYC6	D-galactoside/L-rhamnose binding SUEL lectin protein (*Arabidopsis thaliana*) 43.3%	69.69	24.52	1	3	5	29	3.814E10	155	17.3	8.06
A0A0D3D487	At2g14660 (*Arabidopsis thaliana*) 81.5%	2.14	11.76	1	1	1	1	1.160E10	153	17.3	5.33
A0A0D3D8Z5	D-galactoside/L-rhamnose binding SUEL lectin protein (*Arabidopsis thaliana*) 54.9%	42.06	26.62	1	4	4	13	4.124E9	154	17.3	8.21
A0A0D3DBT3	-	3.62	8.75	2	1	1	1	2.061E7	160	17.4	4.86
A0A0D3DYB7	At1g09245 (*Arabidopsis thaliana*) 41.6%	0.00	8.00	2	1	1	2	2.091E7	150	17.5	9.60
A0A0D3BRQ7	-	6.27	20.13	11	2	2	2	1.974E7	159	17.6	6.05
A0A0D3CU35	Copper transport protein CCH (*Arabidopsis thaliana*) 88.2%	22.01	19.02	2	3	3	7	7.909E7	163	17.7	5.26
A0A0D3AI33	Oleosin (*Brassica rapa* subsp. pekinen..) 57.6%	6.40	7.91	1	2	2	3	1.785E8	177	17.8	9.61
A0A0D3DWR1	Nascent polypeptide-associated complex subunit beta (*Arabidopsis thaliana*) 87.9%	6.69	9.76	3	1	1	2	9.500E6	164	17.8	6.60
A0A0D3AYM3	Non-specific lipid-transfer protein-like protein At5g64080 (*Arabidopsis thaliana*) 71.9%	0.00	8.38	1	1	1	1	2.259E7	179	17.9	6.96
A0A0D3DFT5	Putative uncharacterized protein AT4g24130 (*Arabidopsis thaliana*) 91%	2.86	8.92	1	1	1	1	3.124E7	157	17.9	7.24
A0A0D3CB63	Protease inhibitor/seed storage/lipid transfer protein family protein (*Brassica rapa*) 78.1%	4.09	8.89	1	1	1	1	2.015E7	180	18.1	7.09
A0A0D3AJH8	At5g17340 (*Arabidopsis thaliana*) 73.7%	11.62	9.70	1	1	1	4	7.843E7	165	18.4	7.88
A0A0D3DAL0	Nodulin-related protein 1 (*Arabidopsis thaliana*) 79.9%	2.80	7.91	1	1	1	1	7.340E7	177	18.4	7.77
A0A0D3BF77	Non-specific lipid-transfer protein-like protein At5g64080 (*Arabidopsis thaliana*) 69.4%	21.54	8.02	1	2	2	6	8.504E8	187	18.6	7.65
A0A0D3CX23	At1g72670 (*Arabidopsis thaliana*) 45.1%	15.68	12.12	1	3	3	8	2.698E7	165	18.6	9.14
A0A0D3D5L9	LOB domain-containing protein 4 (*Arabidopsis thaliana*) 96.5%	0.00	8.14	1	1	1	1	7.903E7	172	18.7	7.71
A0A0D3DQ44	Glycine-rich protein GRP18 (*Arabidopsis thaliana*) 52.5%	4.32	10.47	1	2	2	2	8.847E7	191	18.7	9.91
A0A0D3D2Q3	BnaC07g02030D protein (*Brassica napus*) 91.4%	69.24	31.61	1	2	5	18	8.593E8	174	18.7	9.64
A0A0D3E9B3	Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial (*Arabidopsis thaliana*) 92.3%	2.13	7.02	2	1	1	1	4.901E7	171	18.8	8.21
A0A0D3EH62	Isoform 2 of Oleosin GRP-17 (*Arabidopsis thaliana*) 50.4%	29.97	18.28	1	2	2	9	2.815E8	186	18.8	10.90
A0A0D3BT08	Probable glutathione peroxidase 2 (*Arabidopsis thaliana*) 95.9%	23.57	40.24	4	4	4	7	3.857E7	169	18.9	6.70
A0A0D3AUS2	Plant invertase/pectin methylesterase inhibitor domain-containing protein (*Arabidopsis thaliana*) 64%	0.00	4.49	2	1	1	1	2.982E7	178	19.3	9.44
A0A0D3BSW2	Cysteine proteinase inhibitor 2 (*Arabidopsis thaliana*) 71.7%	13.38	13.22	1	3	3	5	2.999E8	174	19.6	10.24
A0A0D3CP94	DNA binding protein (*Brassica campestris*) 92.2%	6.57	5.76	1	1	1	2	9.713E8	191	19.9	11.18
A0A0D3B7X7	Self-incompatibility S1 family protein (*Arabidopsis thaliana*) 44%	20.15	42.78	1	5	5	7	9.391E7	180	20.2	8.02
A0A0D3B5B1	RabGAP/TBC domain-containing protein (*Arabidopsis lyrata*) 70.3%	0.00	7.69	1	1	1	1	1.323E8	182	20.3	7.97
A0A0D3A877	ADP-ribosylation factor C1 (*Arabidopsis thaliana*) 95.1%	0.00	6.01	2	1	1	1	3.303E6	183	20.5	6.64
A0A0D3DKE5	At1g42480 (*Arabidopsis thaliana*) 87.4%	18.78	19.67	1	2	2	6	3.528E7	183	20.7	4.81
A0A0D3D9Q2	T22H22.6 protein (*Arabidopsis thaliana*) 54.6%	85.60	53.23	1	10	10	25	5.445E9	186	21.1	8.47
A0A0D3ATC0	Bifunctional inhibitor/lipid-transfer protein (*Arabidopsis thaliana*) 61%	4.31	7.25	2	1	1	1	2.324E8	207	21.6	8.47
A0A0D3DAP2	At2g03850 (*Arabidopsis thaliana*) 62.1%	21.54	10.45	1	1	1	5	8.082E7	201	21.9	5.69
A0A0D3D125	Dehydrin (*Brassica juncea*) 86.5%	5.97	12.31	4	2	2	2	2.455E7	195	21.9	5.59
A0A0D3B058	Isoform 2 of Oleosin-B6 (*Brassica napus*) 96.7%	65.86	19.14	1	3	3	16	1.019E10	209	22.0	10.13
A0A0D3B8W1	GTP-binding protein SAR1A (*Brassica campestris*) 99%	5.68	13.47	4	2	2	2	9.450E7	193	22.0	7.53

Accession	Homologous protein in Brassicaceae and sequence similarity	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A0D3D9Q3	Plant invertase/pectin methylesterase inhibitor domain-containing protein (*Arabidopsis thaliana*) 53%	11.22	20.00	1	3	3	4	4.547E7	195	22.0	9.26
A0A0D3AJE6	Antimicrobial peptide X precursor (*Stellaria media*) 27.7%	2.30	17.01	1	1	1	2	4.652E7	194	22.2	6.05
A0A0D3A7E4	Ras-related protein RABB1c (*Arabidopsis thaliana*) 99.1%	1.96	5.21	3	1	1	1	4.721E7	211	23.2	7.42
A0A0D3BE38	At1g34360 (*Arabidopsis thaliana*) 70.4%	89.15	19.14	1	6	6	25	1.270E10	209	23.4	9.06
A0A0D2ZUE9	-	13.64	9.39	1	1	1	5	3.611E9	213	23.5	8.78
A0A0D3A8Q1	Ras-related protein RABE1c (*Arabidopsis thaliana*) 98.6%	8.57	11.11	12	2	2	3	5.620E7	216	23.8	7.83
A0A0D2ZRP8	L-type lectin-domain containing receptor kinase IX.1 (*Arabidopsis thaliana*) 77%	0.00	10.67	1	1	1	1	1.820E7	225	25.3	6.54
A0A0D3CXW6	Zinc finger CCCH domain-containing protein 15 (*Arabidopsis thaliana*) 68.2%	2.33	9.05	1	1	1	1		243	26.7	6.42
A0A0D3AW59	DNA-binding storekeeper protein-like protein (*Arabidopsis thaliana*) 56.2%	0.00	3.20	1	1	1	1	8.061E6	250	27.6	9.03
A0A0D3DD15	AT5g24650/K18P6_19 (*Arabidopsis thaliana*) 93.9%	6.08	6.54	1	1	1	2	9.955E6	260	27.7	9.60
A0A0D3B3R2	Putative cysteine-rich repeat secretory protein 17 (*Arabidopsis thaliana*) 50%	5.00	4.72	1	1	1	2	7.018E6	254	28.7	8.38
A0A0D3E433	Ubiquitin-conjugating enzyme E2 28 (*Arabidopsis thaliana*) 89.6%	23.24	19.62	1	3	3	7	5.735E8	265	28.9	8.27
A0A0D3BCP0	Cysteine-rich repeat secretory protein 26 (*Arabidopsis thaliana*) 52.5%	135.38	46.36	1	8	11	39	2.500E9	261	29.1	8.37
A0A0D3D6V2	Ribonuclease 3 (*Arabidopsis thaliana*) 91.4%	36.30	31.62	1	5	5	10	1.425E8	253	29.1	6.73
A0A0D3CHE6	Putative cysteine-rich repeat secretory protein 17 (*Arabidopsis thaliana*) 50.8%	28.33	27.67	1	1	6	8	2.063E8	253	29.1	8.90
A0A0D2ZT39	Cysteine-rich repeat secretory protein 26 (*Arabidopsis thaliana*) 51.9%	23.92	16.41	1	2	5	11	1.641E9	262	29.3	8.75
A0A0D3BSJ8	Putative cysteine-rich repeat secretory protein 37 (*Arabidopsis thaliana*) 58.3%	7.84	11.24	6	1	3	4	1.880E8	258	29.5	8.84
A0A0D3CHE7	Putative cysteine-rich repeat secretory protein 17 (*Arabidopsis thaliana*) 53.8%	74.19	44.88	2	4	11	23	5.508E8	254	29.5	8.41
A0A0D2ZTW0	Putative cysteine-rich repeat secretory protein 37 (*Arabidopsis thaliana*) 57%	5.12	7.78	6	1	3	4	1.778E8	257	29.6	9.09
A0A0D3BE74	Putative cysteine-rich repeat secretory protein 61 (*Arabidopsis thaliana*) 64.6%	20.31	30.50	1	7	7	7	2.266E8	259	29.6	8.92
A0A0D3CHE1	Putative cysteine-rich repeat secretory protein 17 (*Arabidopsis thaliana*) 53.4%	31.39	28.63	1	5	5	10	1.889E8	262	29.7	8.41
A0A0D3BCN5	Cysteine-rich repeat secretory protein 26 (*Arabidopsis thaliana*) 53%	140.55	28.68	1	8	8	40	5.824E8	265	29.7	8.57
A0A0D3CHE2	Putative cysteine-rich repeat secretory protein 17 (*Arabidopsis thaliana*) 50.8%	22.85	10.73	2	1	3	9	4.344E8	261	29.9	8.65
A0A0D3E1Y8	Sec14p-like phosphatidylinositol transfer family protein (*Arabidopsis thaliana*) 90.1%	2.13	3.80	1	1	1	1	1.130E7	263	30.0	8.79
A0A0D3ABP0	Putative cysteine-rich repeat secretory protein 17 (*Arabidopsis thaliana*) 47.4%	31.29	17.60	1	5	5	12	1.050E9	267	31.0	9.17
A0A0D3BW42	Fatty-acid-binding protein 1 (*Arabidopsis thaliana*) 84.6%	10.22	14.44	1	2	3	3	3.716E7	284	31.1	9.04
A0A0D3DUR3	F28C11.15 (*Arabidopsis thaliana*) 55.2%	4.76	9.70	1	1	1	3	1.971E7	268	31.3	9.51
A0A0D3EEX3	Isoform 3 of Binding partner of ACD11 1 (*Arabidopsis thaliana*) 86.9%	3.13	5.72	2	1	1	1	2.351E7	297	31.5	8.47
A0A0D3CCS1	F5O8.22 protein (*Arabidopsis thaliana*) 55.5%	13.39	10.29	1	2	2	4	6.472E7	272	31.6	9.72
A0A0D3DHR7	L-ascorbate peroxidase 3, peroxisomal (*Arabidopsis thaliana*) 90.9%	3.70	4.88	1	1	1	1	6.704E7	287	31.7	7.15
A0A0D3CCR4	F28C11.15 (*Arabidopsis thaliana*) 54.4%	8.26	11.27	1	3	3	3	5.881E7	275	31.8	9.01
A0A0D3BSI6	Cysteine-rich repeat secretory protein 34 (*Arabidopsis thaliana*) 52%	0.00	3.47	1	1	1	1	7.971E6	288	32.7	8.53
A0A0D3D3K5	NAC domain containing protein 36 (*Arabidopsis thaliana*) 81%	2.71	7.37	1	1	1	1	5.787E7	285	32.9	9.36
A0A0D3D4W3	Xyloglucan endotransglucosylase/hydrolase protein 3 (*Arabidopsis thaliana*) 78.5%	4.10	5.86	1	1	1	1	4.970E7	290	33.1	8.10
A0A0D3BLX5	Probable protein phosphatase 2C 59 (*Arabidopsis thaliana*) 95.2%	2.74	7.07	1	1	1	1	6.545E6	311	33.3	4.92
A0A0D3D8I4	NAC domain-containing protein 105 (*Arabidopsis thaliana*) 79.5%	2.58	7.48	1	1	1	1		294	34.0	6.44
A0A0D3D738	F5O8.19 protein (*Arabidopsis thaliana*) 50.8%	5.95	5.05	1	1	1	2	6.947E7	297	34.5	9.55
A0A0D3EH64	GRP21 (*Arabidopsis thaliana*) 36.2%	0.00	3.19	1	1	1	1	1.837E5	345	34.8	9.48
A0A0D3DR34	At1g23580 (*Arabidopsis thaliana*) 57.4%	5.20	7.26	1	2	2	2	4.571E7	303	35.4	9.42
A0A0D3CYZ7	Tetraketide alpha-pyrone reductase 2 (*Arabidopsis thaliana*) 91.3%	2.97	2.49	1	1	1	1	7.376E6	321	35.8	6.99
A0A0D3A1J1	Peroxidase 15 (*Arabidopsis thaliana*) 80.1%	3.17	3.63	2	1	1	1	2.891E7	331	36.1	8.46
A0A0D3CKX5	Leucine-rich repeat protein FLOR1 (*Arabidopsis thaliana*) 64.5%	31.23	18.27	1	5	5	9	3.817E8	323	36.2	9.03
A0A0D3CKX7	Leucine-rich repeat protein FLOR1 (*Arabidopsis thaliana*) 67.1%	23.98	13.46	1	3	3	6	4.135E7	327	36.5	9.25
A0A0D3AVR5	Cytosolic sulfotransferase 12 (*Arabidopsis thaliana*)	2.74	5.57	1	1	1	1		323	36.7	5.69

Accession	Homologous protein in Brassicaceae and sequence similarity	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A0D3DV99	Fatty-acid-binding protein 1 (*Arabidopsis thaliana*) 71.5%	10.42	10.65	1	2	3	4	2.939E7	338	37.4	9.25
A0A0D3BWS5	GDSL esterase/lipase EXL4 (*Arabidopsis thaliana*) 76.2%	17.08	9.38	1	2	3	5	3.141E8	341	37.7	9.91
A0A0D3EGW1	Putative uncharacterized protein At5g08270/T22D6_210 (*Arabidopsis thaliana*) 70.7%	1.68	2.99	1	1	1	1	7.599E7	334	37.7	4.98
A0A0D3AAW1	Mitochondrial transcription termination factor family protein (*Arabidopsis thaliana*) 49.1%	2.53	7.53	1	1	1	1		332	37.8	9.39
A0A0D3DP71	Mediator of RNA polymerase II transcription subunit 32 (*Arabidopsis thaliana*) 82.1%	9.25	8.88	1	2	2	3	2.114E7	349	37.9	5.16
A0A0D3AQX0	GDSL esterase/lipase EXL4 (*Arabidopsis thaliana*) 73%	9.77	8.80	1	1	2	3	3.082E8	341	38.0	9.82
A0A0D2ZRY7	F7F22.16 (*Arabidopsis thaliana*) 28.4%	2.78	4.39	1	1	1	1	1.302E8	342	38.1	8.54
A0A0D3B0B5	Putative uncharacterized protein At5g08540 (*Arabidopsis thaliana*) 74.9%	5.08	5.37	1	1	1	3	6.588E7	354	39.4	5.29
A0A0D3BNK6	Protein disulfide-isomerase like 2-1 (*Arabidopsis thaliana*) 92.2%	1.89	3.05	1	1	1	1	1.490E7	361	39.5	6.32
A0A0D3DBC1	Probable disease resistance protein At5g66910 (*Arabidopsis thaliana*) 55%	2.46	3.57	2	1	1	1	1.261E7	364	41.3	5.54
A0A0D3DEA7	Probable pectate lyase 4 (*Arabidopsis thaliana*) 61.9%	40.89	24.36	1	5	5	12	1.654E8	390	42.2	9.48
A0A0D3EA89	-	2.93	4.91	1	1	1	1	1.864E7	387	42.4	5.25
A0A0D3BJ67	Probable LRR receptor-like protein kinase At1g51890 (*Arabidopsis thaliana*) 68.1%	1.84	5.39	1	1	1	1	5.839E6	371	42.8	6.65
A0A0D3CN47	NADP-dependent alkenal double bond reductase P1 (*Arabidopsis thaliana*) 80%	2.44	6.72	1	1	1	1	1.691E8	387	43.0	8.37
A0A0D3EH63	Oleosin-B3 (*Brassica napus*) 92.1%	5.72	10.32	1	2	2	2	7.004E6	436	43.1	9.51
A0A0D3D248	GDSL esterase/lipase At2g19010 (*Arabidopsis thaliana*) 62.2%	48.84	17.14	2	7	7	14	1.581E8	391	43.8	9.20
A0A0D3DZ85	Exopolygalacturonase clone GBGE184 (*Arabidopsis thaliana*) 80.1%	2.55	3.34	1	1	1	1		419	43.9	8.53
A0A0D3BVN8	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein (*Arabidopsis thaliana*) 78.6%	2.96	2.64	1	1	1	1	5.554E6	379	44.0	9.57
A0A0D3CPJ8	Pectin acetylesterase 2 (*Arabidopsis thaliana*) 70.6%	26.95	18.27	1	6	6	8	4.739E7	416	46.3	7.68
A0A0D3DGG5	Protein phosphatase 2C 57 (*Arabidopsis thaliana*) 78.9%	0.00	2.39	1	1	1	1	4.830E7	419	46.5	7.21
A0A0D2ZSU1	Non-LTR retroelement reverse transcriptase-like protein (*Arabidopsis thaliana*) 31%	0.00	1.42	3	1	1	1	8.873E7	424	46.6	6.99
A0A0D3BDN4	DUF21 domain-containing protein At2g14520 (*Arabidopsis thaliana*) 91.1%	6.53	4.46	1	1	1	2	4.060E7	426	47.5	6.18
A0A0D3EBX2	Nucleosome assembly protein 1;3 (*Arabidopsis thaliana*) 92.3%	0.00	4.12	1	1	1	1	1.273E7	413	47.6	4.58
A0A0D3B3H4	Calcium/calmodulin-regulated receptor-like kinase (*Arabidopsis thaliana*) 89.1%	2.17	5.05	1	1	1	1	6.084E6	436	48.3	8.66
A0A0D3DYH7	Elongation factor 1-alpha 4/3/2/1 (*Arabidopsis thaliana*) 99.3%	2.70	2.45	4	1	1	1	2.799E8	449	49.4	9.11
A0A0D3BEJ4	Spermidine hydroxycinnamoyl transferase (*Arabidopsis thaliana*) 78.7%	6.62	7.10	1	2	2	2	2.070E7	451	50.2	6.14
A0A0D3DUR6	Shaggy-related protein kinase beta (*Arabidopsis thaliana*) 87.1%	5.53	7.13	2	2	2	2	1.893E7	449	51.1	7.49
A0A0D3E0A2	Probable protein phosphatase 2C 52 (*Arabidopsis thaliana*) 91.3%	2.61	5.59	1	1	1	1	3.620E7	465	51.2	5.62
A0A0D3DGR1	Meiotic nuclear division protein 1 homolog (*Arabidopsis thaliana*) 90.9%	2.37	3.85	1	1	1	1	1.717E7	468	53.5	5.34
A0A0D3D2Z5	Scarecrow-like protein 15 (*Arabidopsis thaliana*) 73.5%	0.00	3.85	1	1	1	1	5.787E7	493	54.3	5.90
A0A0D3B8I9	Aspartic proteinase A3 (*Arabidopsis thaliana*) 83.5%	31.66	17.30	2	6	6	11	2.153E8	503	55.0	7.52
A0A0D3CAX2	At2g47010/F14M4.16 (*Arabidopsis thaliana*) 60.3%	0.00	3.21	1	1	1	1	2.429E7	498	56.1	5.62
A0A0D3DH92	Monogalactosyldiacylglycerol synthase 1, chloroplastic (*Arabidopsis thaliana*) 87.5%	14.85	5.83	1	1	1	3	1.626E7	532	58.3	8.84
A0A0D3AER2	Serine carboxypeptidase-like 49 (*Arabidopsis thaliana*) 83.7%	4.98	2.10	1	1	1	2	1.733E7	523	58.4	5.11
A0A0D3D732	F28C11.19 (*Arabidopsis thaliana*) 72.8%	17.88	8.48	1	2	3	6	8.149E7	507	58.9	9.42
A0A0D3BUX9	HSC70-1 (*Arabidopsis lyrata* subsp. ly..) 91.3%	11.81	7.22	11	1	2	4	2.698E7	540	59.2	5.11
A0A0D3C4D9	Chaperone protein dnaJ 13 (*Arabidopsis thaliana*) 86.1%	0.00	4.99	1	1	1	1	2.963E7	541	59.6	8.98
A0A0D3AB88	AGC (CAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein (*Arabidopsis thaliana*) 55.1%	1.98	2.45	5	1	1	1	8.854E6	530	60.3	8.44
A0A0D3DP31	MLO-like protein 2 (*Arabidopsis thaliana*) 72.9%	1.61	2.25	1	1	1	1	7.917E6	534	61.0	9.14
A0A0D3DR35	F28C11.19 (*Arabidopsis thaliana*) 79.8%	19.14	5.59	1	1	2	6	1.078E8	537	61.2	6.90
A0A0D3CLG3	Serine carboxypeptidase-like 49 (*Arabidopsis thaliana*) 81%	4.73	7.80	2	3	3	4	5.804E8	564	62.9	5.87

Accession	Homologous protein in Brassicaceae and sequence similarity	Score	Coverage	Proteins	Unique Peptides	Peptides	PSMs	Area	AAs	MW [kDa]	pI
A0A0D3EFT1	Auxin-responsive GH3 family protein (*Arabidopsis thaliana*) 80.1%	3.04	2.19	2	1	1	1	3.874E6	594	67.1	7.84
A0A0D3CZF0	Protein NRT1/ PTR FAMILY 2.13 (*Arabidopsis thaliana*) 85.5%	0.00	1.62	1	1	1	2	3.564E6	619	68.4	8.82
A0A0D3BR77	Arginine biosynthesis bifunctional protein ArgJ, chloroplastic (*Arabidopsis thaliana*) 92.2%	2.34	3.32	1	1	1	1	1.630E9	662	70.1	7.34
A0A0D3C3D3	Heat shock cognate protein HSC70 (*Brassica napus*) 99.4%	17.39	6.04	7	1	2	5	2.715E7	646	70.8	5.19
A0A0D3CW61	Long chain acyl-CoA synthetase 9, chloroplastic (*Arabidopsis thaliana*) 91.6%	16.58	5.19	1	3	3	7	2.246E8	693	75.9	6.67
A0A0D3E748	-	2.22	1.34	1	1	1	1	1.204E8	672	76.0	8.43
A0A0D3DIY4	Receptor like protein 54 (*Arabidopsis thaliana*) 71.1%	2.24	2.28	1	1	1	1	3.655E6	790	87.7	6.43
A0A0D3DTR6	-	0.00	2.51	1	1	1	1	9.515E7	797	89.3	5.35
A0A0D3BHM3	Protein TOC75-3, chloroplastic (*Arabidopsis thaliana*) 94.1%	2.30	1.83	1	1	1	1	2.515E7	820	89.4	8.53
A0A0D3B5T9	ATH subfamily protein ATH8 (*Arabidopsis thaliana*) 85%	2.52	2.80	1	1	1	1		822	91.5	9.33
A0A0D3AAR5	Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g61610 (*Arabidopsis thaliana*) 77.6%	2.42	2.17	1	1	1	2	1.699E8	830	94.0	7.91
A0A0D3BFH0	Transcription factor GTE10 (*Arabidopsis thaliana*) 64.3%	3.13	3.16	1	1	1	1		855	95.9	4.98
A0A0D3AM31	Cytochrome b561, DM13 and DOMON domain-containing protein At5g54830 (*Arabidopsis thaliana*) 91.2%	1.66	1.54	1	1	1	1		911	101.0	5.88
A0A0D3E250	Calmodulin-binding transcription activator 2 (*Arabidopsis thaliana*) 82.3%	0.00	0.80	1	1	1	1		995	111.0	6.18
A0A0D3CIC9	Pre-mRNA-processing protein 40B (*Arabidopsis thaliana*) 76.5%	0.00	1.57	1	1	1	1	2.471E7	1020	116.8	7.24
A0A0D3A401	Phytochrome E (*Arabidopsis thaliana*) 88%	4.61	0.63	1	1	1	2	4.816E7	1116	122.6	6.73
A0A0D3DHG1	Cell division cycle 20.1, cofactor of APC complex (*Arabidopsis thaliana*) 94.7%	13.93	5.82	2	4	4	5	4.277E7	1186	132.2	8.78
A0A0D3ARN3	Putative uncharacterized protein At1g79190 (*Arabidopsis thaliana*) 85.3%	2.40	1.30	1	1	1	1	2.713E6	1310	144.4	6.46
A0A0D2ZSI1	AT hook motif-containing protein, putative (*Oryza sativa*) 45.2%	0.00	2.09	1	1	1	1		1293	147.0	6.83
A0A0D3DUF8	Integrin-linked protein kinase-like protein (*Arabidopsis thaliana*) 79.9%	15.31	1.90	1	1	3	5	4.284E7	1688	185.3	6.77
A0A0D3BSR4	Myosin-1 (*Gossypium arboreum*) 47%	2.35	1.03	1	1	1	1	1.749E7	1756	196.2	4.91
A0A0D3AI34	-	37.07	5.95	1	6	6	9	6.253E7	2352	234.9	10.32
A0A0D3E741	Protein CHROMATIN REMODELING 4 (*Arabidopsis thaliana*) 78.3%	0.00	1.10	1	1	1	1	1.485E7	2271	253.6	6.81